A FEW-SAMPLE STRATEGY FOR GUITAR TABLATURE TRANSCRIPTION BASED ON INHARMONICITY ANALYSIS AND PLAYABILITY CONSTRAINTS

Grigoris Bastas[1,2] *Stefanos Koutoupis*[2] *Maximos Kaliakatsos-Papakostas*[1]
Vassilis Katsouros[1] *Petros Maragos*[2]

[1]Institute for Language and Speech Processing, Athena R.C., Athens, Greece
[2]School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

ABSTRACT

The prominent strategical approaches regarding the problem of guitar tablature transcription rely either on fingering patterns encoding or on the extraction of string-related audio features. The current work combines the two aforementioned strategies in an explicit manner by employing two discrete components for string-fret classification. It extends older few-sample modeling strategies by introducing various adaptation schemes for the first stage of audio processing, taking advantage of the inharmonic characteristics of guitar sound. Physical limitations and common standards of human performers are incorporated in a genetic algorithm which constitutes a second contextual-based module that further processes the initial audio-based predictions. The proposed methods are evaluated on both annotated guitar performances and isolated note recordings.

Index Terms— few-sample strategy, inharmonicity, genetic algorithm

1. INTRODUCTION

Automatic Music Transcription is the problem of extracting formal notation from the music signal. While music scores have been the prominent notational form for western music, in the case of guitar, tablature-notation has gained much popularity in the last decades, especially among novice and self-taught guitar players. Guitar tablature represents musical performances as sequences of string-fret combinations. The typical music score leaves out this information. Due to the guitar's design, same pitch notes can be played in more than one position on the fretboard, a feature that renders tablature transcription a challenging task.

Several research teams have worked on playable guitar tablature generation by taking advantage of contextual information given a series of note events in symbolic notation (i.e. music scores or MIDI), by employing graph representations and dynamic programming [1, 2, 3] or optimization algorithms [4, 5, 3]. Others have worked on audio signal analysis, mainly for polyphonic performances, still relying mostly on playability constraints and multi-pitch estimation [6, 7, 8]. Other works have focused on developing audio-based accurate string classification models by exploiting the signal properties in relevance to the strings' physical characteristics, such as inharmonicity, facing note instances (or chords) as independent events in time. Some of these methods require several annotated note instances for training and thus remain restricted to the guitars involved in the employed training set [9, 10, 11, 12]. Neural networks applied on audio spectral features suffer from the same limitations [13, 14, 15]. It is also unclear whether the latter manage to implicitly encode both string-related audio properties and sequential information. Finally, some audio-based approaches have introduced more agile adaptation strategies relying on just a few samples drawn from one fret per string [16, 17].

This work is motivated by the claim that the articulated notes in a performance cannot be seen as independent events. Our main contribution is that audio-based and context-based processing are explicitly combined in a successive and complementary setting. The first stage of string-fret classification relies on the extraction of the inharmonicty coefficient (β) from each note instance. Onsets and pitches are considered known. For the contextual-based stage of processing, we incorporated physical constraints of guitar playing to a genetic algorithm (GA) that encodes the most possible string-fret transitions. This second stage of string classification provides the final tablature.

A phase of adaptation is required for the audio-based classification in order to acquire first estimates of each string's inharmonic behavior, while no training is required for the GA. We regard our string detector as a potentially applicable component of an easy to train tablature transcription system that processes undistorted guitar signals. Such a system can be deployed in a realistic scenario where a guitar player provides a small amount of note samples in order to "tune" the system on a specific guitar, rendering it practicable for personal use. We study several novel few-sample adaptation settings with vary-

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (NLP-Theatre, project code: T1EDK-00508).

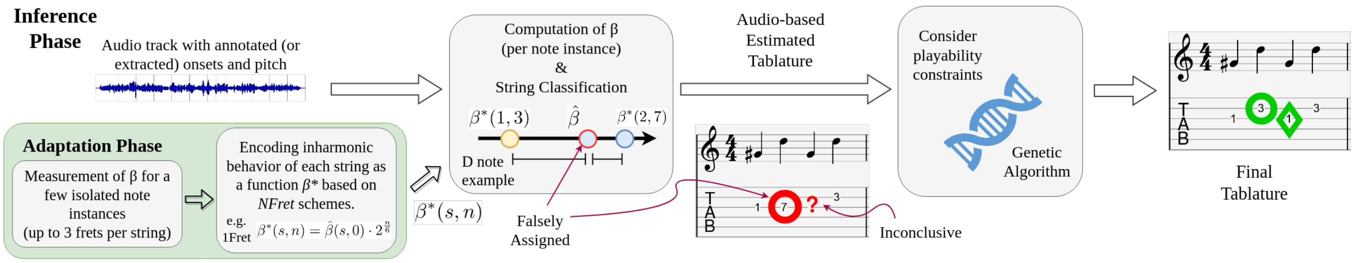

Fig. 1. Flow diagram of the proposed method.

ing impact on the accuracy of the audio-based string classifier and varying effort required from the user for the sample collection.

2. METHOD DESCRIPTION

The proposed method is presented in Fig. 1. First, we need to adapt the audio-based detector to the specific instrument which is to be used. The inharmonicity coefficient β is computed only for a small amount of recorded note instances (from 1 to 3 frets for all 6 strings) using partial tracking. We generalize our estimations for each fret n and string s, based on the measurements ($\hat{\beta}$) from these few samples. This way, we acquire a function $\beta^*(s,n)$ that describes the expected inharmonic behavior of each string s across the fretboard. For the inference phase, we similarly compute the inharmonicity coefficients ($\hat{\beta}$) for each detected note instance in a recorded performance and assign to it the string label with the closest β^*. This rudimentary tablature estimation is subsequently fed to a genetic algorithm which assigns string labels to inconclusive note instances (see Section 2.1) and modifies certain predictions to improve overall results.

2.1. Inharmonicity Coefficient Computation

An instrumental timbre can be analyzed as a superposition of sinusoids with varying amplitudes and frequencies. These (partial) frequencies are considered to be harmonic when they are placed ideally at the fundamental frequency's (f_0) integral multiplies. On many stringed instruments though, such as guitar and piano, a notable deviation of the partials from the harmonics can be observed [18], mainly due to the stiffness of the strings [19]. The kth overtone can be estimated by:

$$f_k = k \cdot f_0 \cdot \sqrt{1 + \beta \cdot k^2}. \tag{1}$$

Inharmonicity coefficient β differs also among frets n of the same string s following the relation below [16]:

$$\beta(s,n) = \beta(s,0) \cdot 2^{\frac{n}{6}}. \tag{2}$$

For the purpose of inharmonicity coefficient computation, we employ a variation of the algorithm proposed in Barbancho et al. [16] for monophonic performances. Our method follows the aforementioned approach in the first stage of β measurement with partial tracking, using shifted frequency windows of $f_0/2$ width. Yet, we do not restrict ourselves to measurments from the open string instances, and instead we measure β for whichever fret needed, since we regard this value as a feature for classification. We apply the FFT algorithm on 60ms audio segments, which corresponds to 16th notes played on 250bpm, an adequate window for most guitar performances in general. We account only for the first 30 partials. When the inharmonicity coefficient computation algorithm does not provide valid results (i.e. very small or large computed $\hat{\beta}$ values relative to the estimated β^*), we mark those note instances as inconclusive.

2.2. Audio-Based String Classifier

As a first step towards accurate string classification, a straightforward method is proposed, based on the measured inharmonicity coefficient. The note instance under inspection with measured coefficient $\hat{\beta}$ is assigned to the string-fret label with the closest estimation β^*. Preliminary testing has revealed that a more nuanced classification approach involving maximum a posteriori probability (MAP) estimates (as in Hjerrild and Christensen [17]), did not work well in the context of realistic guitar performances. Furthermore, our method permits the use of even a single adaptation sample per string. Michelson et al. [12] manage to employ a similar Bayesian classifier, with the cost of relying on many annotated data for training. It is for these reasons that we proposed a more straight-forward classification method (i.e. Euclidean distance criterion).

In order to apply the classifier for inference, an adaptation phase is preceded to extract the estimated values β^*. We suggest four schemes for adaptation. For the most basic scheme that we call 1Fret method, we measure the $\beta(s,0)$ coefficient of every open string and estimate the $\beta(s,n)$ associated to each fret n with equation (2). This method is in the same line with the adaptation settings proposed in [16] or [17], and permits the fastest sample collection in a realistic scenario, since it requires one (or just a few) samples for each string.

Instrument-specific irregularities like neck warping are

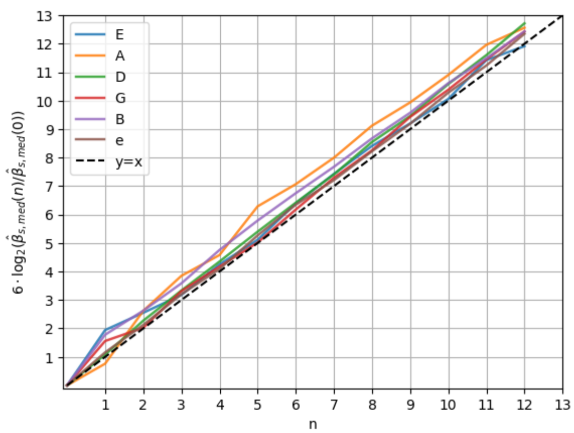

Fig. 2. Irregularity of inharmonic behavior for each string.

common in guitars, so we assume that, in some cases, as the hand moves towards the body of the instrument, equation (2) may not hold as strong. Our assumption is supported by measurements on the GuitarSet dataset [20]. We computed the median inharmonicity coefficients $\hat{\beta}_{med}$ of all note instances for the first 12 frets. Choosing a string s, by calculating $6 \cdot \log_2(\hat{\beta}_{s,med}(n)/\hat{\beta}_{s,med}(0))$ and plotting the results for each fret $n \in \{0, ..., 12\}$, we would expect an approximation of a $y = x$ curve, based on equation (2). However, in the case of some strings, there occur slight but notable deviations from the expected results (see Fig. 2).

As follows, we suggest a more general version of (2) by replacing n with the linear expression $a \cdot n + b$. For the proposed 3Fret method, we can compute a and b after the measurement of the inharmonicity coefficient from two frets (i, j) of our choice and an open string (0th fret) by solving the system below for a and b:

$$\begin{cases} \hat{\beta}(s,i) = \hat{\beta}(s,0) \cdot 2^{\frac{a \cdot i + b}{6}} \\ \hat{\beta}(s,j) = \hat{\beta}(s,0) \cdot 2^{\frac{a \cdot j + b}{6}}. \end{cases} \tag{3}$$

Estimates $\beta^*(s, n)$ can then be acquired for all frets using the suggested generalized version of (2). We devised two more methods that require, for each string, the computation of $\hat{\beta}$ from one fret and the open string. These extra schemes called 2FretA and 2FretB emerge from our previous analysis if we assume $b = 0$ or $a = 1$ respectively.

2.3. Contextual-based Classification

In our analysis we aim to capitalize on the physical constraints imposed by the guitar fretboard layout to achieve better results. Tablatures played by human performers hold a certain structure that is convenient and logical. For instance, large fret-wise distances are counter-intuitive when not necessary. To that end, we employ a genetic algorithm that favors results from the classifier's output while accounting for the playability of the tablature. This way, the genetic algorithm can simultaneously resolve unclassified note instances, and correct evidently wrong ones (see Fig. 1). We model this task as an optimization problem where a fitness function is minimized:

$$\arg\min_{\mathbf{x} \in T}(g(\mathbf{x}) - 2 \cdot h(\mathbf{x}, \mathbf{x_0})), \tag{4}$$

where g represents a function that encodes the playability of a tablature $\mathbf{x}$ of an entire piece (i.e. a sequence of vectors $(s_t, n_t) \in \{1, .., 6\} \times \{0, ..., 22\}$, with t indicating the note position index within the sequence) and h encodes the similarity of the output with the audio-based prediction $\mathbf{x_0}$, i.e. the rate of common (s_t, n_t) vectors. T constitutes the search space, that is all possible tablature layouts that realize the pitches of the piece. A pool of 40,000 individuals (i.e. random variations of tablature $\mathbf{x_0}$ with resolved inconclusive notes) is evolved with elitist selection, employing tournament parent selection of size 5, a typical two-point random crossover function and mutation. When individuals selected for mutation (with probability 0.2), each of the string-fret combinations (s_t, n_t) are altered (with probability 0.1) given pitch equivalent values.

Function g constitutes the sum of 5 different components normalized by the length of $\mathbf{x}$. Ideas concerning lateral hand movement were drawn from [4] and were extended by considering vertical movement as well. Namely, these components are: I) the number of times a different fret is pressed, II) the sum of Euclidean (string-fret) distances between each note and the average position of the 6 neighboring note instances that do not span more than a time threshold of 1 sec, III) the sum of fret-wise distances, IV) the sum of string-wise distances, V) an open string reward.

3. DATASETS

The main dataset used for evaluation consists of 52 performances picked out of 360 in total, contained in the GuitarSet dataset [20]. Since we only accounted for monophonic performances, we selected the audio tracks where all subsequent notes had at least a gap of 60ms between their onsets. GuitarSet includes 4 different versions of each track: the first is recorded with a microphone, one with 6-channel recordings using a hexaphonic pickup, one where the 6-channel tracks are debleeded in order to reduce noise and artefacts, and one where the 6 channels are mixed into one. We only made use of the one-channel versions. We also tested our audio-based string detector on the dataset presented by Hjerrild and Christensen [17] which contains recordings of isolated note instances for all string-fret combinations up to the 12th fret, for two guitars, an electric (Les Paul Firebrand) and an acoustic (Martin DR). Every fret is recorded 9 times for the Firebrand and 10 for the Martin guitar.

Adaptation Method	Martin	Firebrand
3Fret	99.9%	97.7%
2FretA	99.9%	97.7%
2FretB	99.9%	96.5%
1Fret	94.6%	97.5%
MAP-optimal [17]	100%	97.1%

Table 1. Accuracy measures of audio-based classification on the dataset introduced in [17].

4. EVALUATION

4.1. Testing on Isolated Note Recordings

Firstly, we tested our method on the dataset provided by Hjerrild and Christensen [17]. We computed the medians of the inharmonicity coefficients of all isolated instances available from the 0th (open string), 3rd and 12th fret, according to the employed adaptation *NFret* scheme. Since only the MIDI notes are provided in the annotations, a basic f_0 estimation algorithm was employed for this experiment using peak estimation around the expected frequency. In Table 1, we present accuracy measures (i.e. correct predictions over all instances including inconclusive ones) on the frets which were not used for adaptation in each case. The 3Fret and 2FretA methods exhibited the same performance outperforming the standard 1Fret scheme for both the acoustic (Martin) and electric guitar (Firebrand). For those two schemes, the results are almost identical (for Martin) or even better (for Firebrand) compared to the MAP-optimal classifier proposed by Hjerrild and Christensen [17]. The rate of inconclusive instances for our model was 0% for Martin and 1.1% for Firebrand, for all schemes.

4.2. String Detection in Guitar Performances

For the evaluation of our method on a realistic context of recorded performances, we used the four few-sample schemes proposed above for the audio-based string classification, and a genetic algorithm to account for playability constrains. We ran tests for both microphone and pickup recordings of the monophonic subset that we drew from GuitarSet. For the initial estimations of β from the adaptation phase, we chose 5 relatively clear recorded samples of all 6 strings for the 0th, 3rd and 12th fret. In a personalized adaptation scenario where the samples wouldn't have been picked from performances, even one clear recording would be enough. In our case though, in order to avoid biasing, the median value of β was chosen for each of the above string-fret combinations.

As presented in Table 2, the 4 adaptation schemes raise accuracy values which range from 82.2% to 85.1%. Pickup exhibits slightly better results compared to the microphone recordings, which is normal since more artefacts occur in the latter case. We measured 2.3% inconclusive rate for pickup and 2.6% for microphone. GA improves our results in every case, adding up from 7% to 8.9%, with the 1Fret method exhibiting the lower accuracy rates, as expected. It is clear that, for each guitar, a high accuracy output of the audio-based classifier favors the genetic algorithm's results. In this experiment, the 2FretB and 2FretA adaptation schemes led to the best overall results for the pickup and microphone recordings, respectively. The code for the experiments can be found online[1].

Adaptation Method	Audio Classification Accuracy	GA Classification Accuracy
Pickup		
3Fret	84.4%	91.8%
2FretA	84.7%	91.6%
2FretB	**85.1**%	**92.9**%
1Fret	83.2%	90.8%
Microphone		
3Fret	83.3%	92.1%
2FretA	83.6%	**92.3**%
2FretB	**84.0**%	92.2%
1Fret	82.2%	91.1%

Table 2. Accuracy of both classification stages on the monophonic performances of the GuitarSet dataset.

5. CONCLUSIONS AND FUTURE DIRECTIONS

In this work, we presented a system for string-fret classification implementing few-sample adaptation strategies of varying complexity. Our results indicate that our audio-based classifier is on the right track for the generation of accurate guitar tablatures, since it raises comparable results to another novel and established method [17], and can perform well on realistic audio recordings. It was verified that the genetic algorithms can be leveraged to provide substantial improvement, by encapsulating playability constraints. Hence, we consider our main novelties to be: I) a few-sample approach involving various adaptation schemes applicable to realistic performances with the ability to manage guitar's common physical irregularities, II) the combination of an audio signal processing approach with the analysis of symbolic representations in a sequential context. A future challenge would be to generalize the proposed method for polyphonic performances and even study specific guitar techniques, like bending, which can affect the string's inharmonicity parameters. Interestingly, polyphony and bends don't occur very often in bass guitar performances. So, we would expect that with adjustments that address the idiosyncratic techniques of this instrument (e.g. slapping), the current method would be quite apt in this context. This task is also left as future work.

[1] `https://github.com/estafons/inh-ga-tabs.git`

6. REFERENCES

[1] Aleksander Radisavljevic and Peter F Driessen, "Path difference learning for guitar fingering problem," in *ICMC*, 2004, vol. 28.

[2] Masanobu Miura, Isao Hirota, Nobuhiko Hama, and Masazo Yanagida, "Constructing a system for finger-position determination and tablature generation for playing melodies on guitars," *Systems and Computers in Japan*, vol. 35, no. 6, pp. 10–19, 2004.

[3] Joao Victor Ramos, André Stylianos Ramos, Carlos N Silla, and Danilo Sipoli Sanches, "An evaluation of different evolutionary approaches applied in the process of automatic transcription of music scores into tablatures," in *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2016, pp. 663–669.

[4] Daniel R Tuohy and Walter D Potter, "A genetic algorithm for the automatic generation of playable guitar tablature," in *ICMC*, 2005, pp. 499–502.

[5] Joao Victor Ramos, Andre Stylianos Ramos, Carlos N Silla, and Danilo Sipoli Sanches, "Comparative study of genetic algorithm and ant colony optimization algorithm performances for the task of guitar tablature transcription," in *2015 Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, 2015, pp. 228–233.

[6] Ana M Barbancho, Anssi Klapuri, Lorenzo J Tardón, and Isabel Barbancho, "Automatic transcription of guitar chords and fingering from audio," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 915–921, 2011.

[7] Kazuki Yazawa, Daichi Sakaue, Kohei Nagira, Katsutoshi Itoyama, and Hiroshi G Okuno, "Audio-based guitar tablature transcription using multipitch analysis and playability constraints," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 196–200.

[8] Gregory Burlet and Ichiro Fujinaga, "Robotaba guitar tablature transcription framework.," in *ISMIR*, 2013, pp. 517–522.

[9] Jakob Abeßer, "Automatic string detection for bass guitar and electric guitar," in *International Symposium on Computer Music Modeling and Retrieval*. Springer, 2012, pp. 333–352.

[10] Christian Dittmar, Andreas Männchen, and Jakob Abeber, "Real-time guitar string detection for music education software," in *2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*. IEEE, 2013, pp. 1–4.

[11] Christian Kehling, Jakob Abeßer, Christian Dittmar, and Gerald Schuller, "Automatic tablature transcription of electric guitar recordings by estimation of score-and instrument-related parameters.," in *DAFx*, 2014, pp. 219–226.

[12] Jonathan Michelson, Richard Stern, and Thomas Sullivan, "Automatic guitar tablature transcription from audio using inharmonicity regression and bayesian classification," in *Audio Engineering Society Convention 145*. Audio Engineering Society, 2018.

[13] Thierry Gagnon, Steeve Larouche, and Roch Lefebvre, "A neural network approach for preclassification in musical chords recognition," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. IEEE, 2003, vol. 2, pp. 2106–2109.

[14] Eric J Humphrey and Juan P Bello, "From music audio to chord tablature: Teaching deep convolutional networks toplay guitar," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014, pp. 6974–6978.

[15] Andrew Wiggins and Youngmoo Kim, "Guitar tablature estimation with a convolutional neural network.," in *ISMIR*, 2019, pp. 284–291.

[16] Isabel Barbancho, Lorenzo J Tardon, Simone Sammartino, and Ana M Barbancho, "Inharmonicity-based method for the automatic generation of guitar tablature," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1857–1868, 2012.

[17] Jacob Møller Hjerrild and Mads Græsbøll Christensen, "Estimation of guitar string, fret and plucking position using parametric pitch estimation," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 151–155.

[18] Harvey Fletcher, E Donnell Blackham, and Richard Stratton, "Quality of piano tones," *The Journal of the Acoustical Society of America*, vol. 34, no. 6, pp. 749–761, 1962.

[19] Jacob Møller Hjerrild, Silvin Willemsen, and Mads Græsbøll Christensen, "Physical models for fast estimation of guitar string, fret and plucking position," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2019, pp. 155–159.

[20] Qingyang Xi, Rachel M Bittner, Johan Pauwels, Xuzhou Ye, and Juan Pablo Bello, "Guitarset: A dataset for guitar transcription.," in *ISMIR*, 2018, pp. 453–460.