#2491

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

22-27 MAY 2022
ICASSP 2022
SINGAPORE

Motivation and Challenges

- Guitar tablature represents musical parts as sequences of string ($s \in \{1, ..., 6\}$) and fret ($n \in \{0, ..., 22\}$) combinations. See:

$$\mathbf{x} : \{0, ..., N\} \to \{1, .., 6\} \times \{0, ..., 22\}, \quad (1)$$

  where N is the number of note instances (to be) played.

- Tablature is an alternative notational form to music score. Scores do not contain string-fret information. Tablature is nowdays very common among self-taught and novice guitarists.
- Tablature automatic transcription is a demanding task, because, except for pith information, it requires accurate string detection/classification.
- String detection is challenging because same pitch/notes can be articulated in different fretboard positions.

Situating Ourselves

Various Approaches for Automatic Tablature Transcription:

- Pitch-Based **Playability Approach** capitalizing on playbility constraints
- String-Specific **Audio Approach** capitalizing on audio information extraction
- Special Case: Latent Information Approach – Neural Networks on Generic Audio Spectral Features

We draw upon previous works on:

- partial detection and inharmonicity analysis (Audio Approach)
  - few-sample adaptation strategies for string classification model [1, 2]
- genetic algorithms (GA) for playability constraints encoding (Playability Approach) [3]

Our main Contributions:

- Explicit combination of Audio and Playability Approaches for accurate transcription
- Introduction of various few-sample adaptation schemes for inharmonicity-based audio string detection

Restricted ourselves to **monophonic performances**

Inharmonicity

An ideal string produces sound waves with harmonic partials (i.e. integral multiples of fundamental frequency f_0):

$$f_k = k \cdot f_0 \quad (2)$$

Actual guitar strings produce inharmonic sound:

$$f_k = k \cdot f_0 \cdot \sqrt{1 + \beta \cdot k^2} \quad (3)$$

Inharmonicity coefficient (β) in relation to string (s) and fret (n):

$$\beta(s, n) = \beta(s, 0) \cdot 2^{\frac{n}{6}} \quad (4)$$

Inharmonicity coefficient measurement/computation:

- FFT algorithm on 60ms audio segments, starting from onset timestamp
- 30 partials taken into account
- partial tracking using shifted frequency windows of $f_0/2$ width, with gradual window centering corrections based on the iterative method for β estimates' extraction suggested in [1]

Method Description

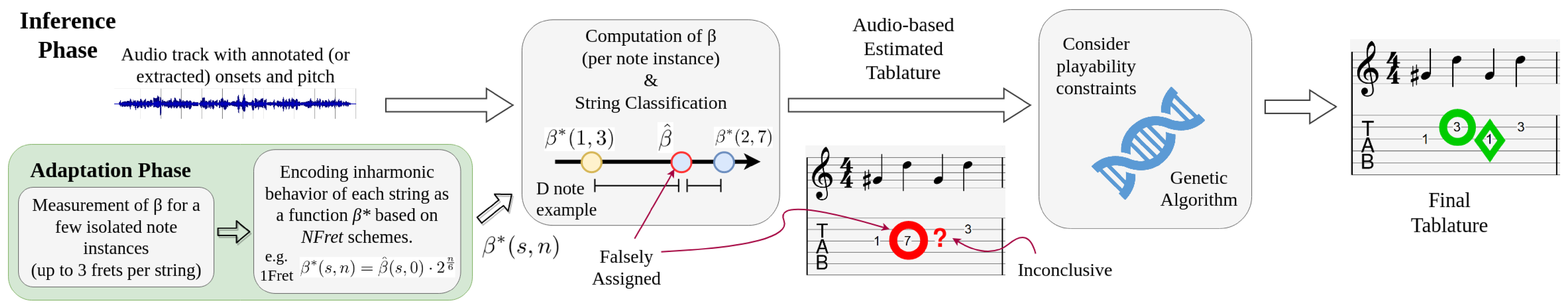

Figure 1: Flow diagram of the proposed method.

1. Few-sample adaptation, that is extraction of estimates β^* for the whole fretboard relying on inharmonicity coefficient measurements ($\hat{\beta}$) from a small subset of possible string-fret combinations using a generalized version of 4:

$$\beta^*(s, n) = \hat{\beta}(s, 0) \cdot 2^{\frac{a \cdot n + b}{6}} \quad (5)$$

   where a, b are found:

   - in the simple **1Fret** scheme by setting $a = 1$ and $b = 0$, relying only on open string samples
   - in the most complete **3Fret** scheme, where we additionally consider frets i and j by solving:

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

   - in the intermediate **2FretA** and **2FretB** schemes, considering only one extra fret i and setting $b = 0$ or $a = 1$, respectively.

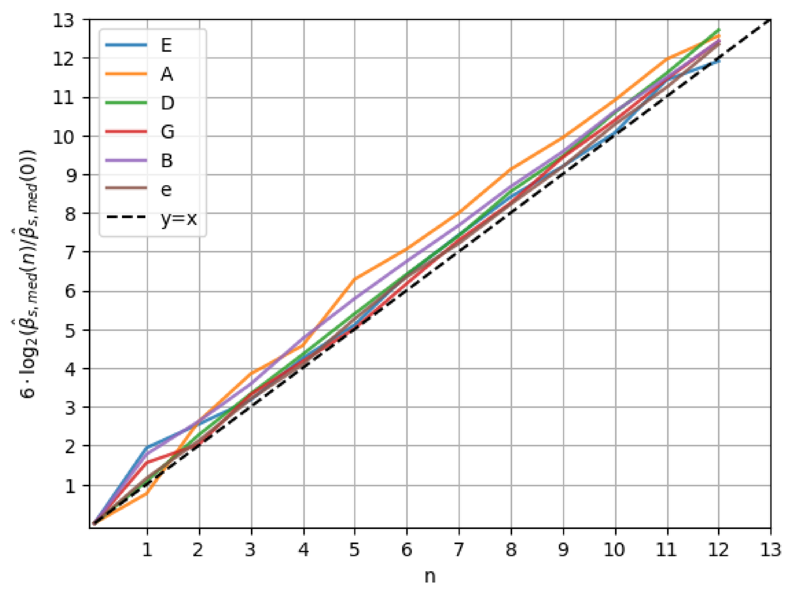

Figure 2: Irregularity of inharmonic behavior for each string.

- Instrument-specific irregularities like neck warping are common in guitars, so we assume that, in some cases, as the hand moves towards the body of the instrument, equation (4) may not hold as strong
- the assumption is supported by measurements on the GuitarSet dataset [4]. 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 4
- there occur slight but notable deviations from the expected results (see Fig.1)

2. Audio-Based String Classification:

   - beta computation for each note instance of the incoming audio track, with onsets and pitches being extracted.
   - measure euclidean distance between the $\hat{\beta}$ and the estimates (β^*) of all the possible same-pitch string-fret combinations, in order to detect the active string

3. Contextual-Based Classification: We model this task as an optimization problem where a fitness function is minimized:

$$\underset{\mathbf{x} \in T}{\arg\min}(g(\mathbf{x}) - 2 \cdot h(\mathbf{x}, \mathbf{x_0})), \quad (7)$$

   - 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
   - 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
   - pool of 40,000 individuals (i.e. random variations $\mathbf{x}$ of $\mathbf{x_0}$ with resolved inconclusive notes) is evolved with elitist selection, employing:
     - tournament parent selection of size 5
     - typical two-point random cross-over function
     - mutation: when individuals are chosen for mutation (with probability 0.2) each of the string-fret combinations (s_t, n_t) is altered (with probability 0.1) given pitch equivalent values

Evaluation

First Experiment:

- results comparable to well-established method – NFret adaptation schemes enable better recognition performance

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 [2]	100%	97.1%

Table 1: First Experiment: accuracy measures of audio-based classification on the dataset introduced in [2].

Second Experiment:

- GA: substantial improvement over initial audio predictions

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: Second Experiment: accuracy of both classification stages on the monophonic performances of the GuitarSet dataset.

Conclusions and Future Directions

Conclusions:

- proposed strategy for pitch-based and string-specific approaches: robust in realistic monophonic performances

Future Work:

- generalize method for polyphonic performances
- study specific guitar techniques (bending, etc.) and adjust to bass guitar

References

[1] Isabel Barbancho et al. "Inharmonicity-based method for the automatic generation of guitar tablature". In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.6 (2012), pp. 1857–1868.

[2] 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.

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

[4] Qingyang Xi et al. "GuitarSet: A Dataset for Guitar Transcription." In: *ISMIR*. 2018, pp. 453–460.

LaTeX TikZposter