AUTOMATIC TABLATURE TRANSCRIPTION OF ELECTRIC GUITAR RECORDINGS BY ESTIMATION OF SCORE- AND INSTRUMENT-RELATED PARAMETERS

Christian Kehling
Fraunhofer IDMT,
Ilmenau, Germany

*Jakob Abeßer**
Fraunhofer IDMT,
Ilmenau, Germany
jakob.abesser@idmt.fraunhofer.de

Christian Dittmar
Fraunhofer IDMT,
Ilmenau, Germany

Gerald Schuller
Technical University of Ilmenau,
Ilmenau, Germany

ABSTRACT

In this paper we present a novel algorithm for automatic analysis, transcription, and parameter extraction from isolated polyphonic guitar recordings. In addition to general score-related information such as note onset, duration, and pitch, instrument-specific information such as the plucked string, the applied plucking and expression styles are retrieved automatically. For this purpose, we adapted several state-of-the-art approaches for onset and offset detection, multipitch estimation, string estimation, feature extraction, and multi-class classification. Furthermore we investigated a robust partial tracking algorithm with respect to inharmonicity, an extensive extraction of novel and known audio features as well as the exploitation of instrument-based knowledge in the form of plausability filtering to obtain more reliable prediction. Our system achieved very high accuracy values of 98 % for onset and offset detection as well as multipitch estimation. For the instrument-related parameters, the proposed algorithm also showed very good performance with accuracy values of 82 % for the string number, 93 % for the plucking style, and 83 % for the expression style.

Index Terms - playing techniques, plucking style, expression style, multiple fundamental frequency estimation, string classification, fretboard position, fingering, electric guitar, inharmonicity coefficient, tablature

1. INTRODUCTION

Audio recordings of plucked string instruments can be described as a sequence of acoustic note events having a characteristic harmonic structure, which strongly depends on the type of instrument and the playing techniques are being used. Scores and tablatures are common notation formats to store the most important parameters to describe each played note. In order to automatically generate such notations from a recorded audio signal, these parameters must be estimated beforehand.

As will be detailed in Section 3, various publications in the field of Music Information Retrieval (MIR) focused on the automatic extraction of either score-related parameters such as onset, offset, and pitch (a tasks that is commonly referred to as automatic music transcription), or instrument-related parameters such as the applied playing techniques and fretboard positions on string instruments. This work expands these approaches by fusing single parameter estimation algorithms to an overall transcription framework, which is tailored towards instrument-specific properties of the electric guitar.

The proposed automatic transcription algorithm extracts essential information about the recorded music piece that allows comparison with a ground truth notation. Hence possible application scenarios are music education software such as Songs2See[1] and BandFuse[2] as well as music games such as RockSmith[3]. Furthermore, the transcription algorithm can be applied for a detailed expressive performance analysis that provides information about artist-specific peculiarities related to micro-timing or to the preferred playing techniques. In combination with a sound synthesis algorithm, an efficient parametric audio coding model with very low bit rates can be realized due to the very compact symbolic representation of the instrument recording.

The paper is structured as follows. First, Section 2 provides important basics of the guitar sound generation. After a review of the related work in Section 3, we explain the proposed transcription algorithm in detail in Section 4. Finally, Section 5 describes all evaluation experiments and Section 6 summarizes this work.

2. GUITAR SPECIFIC BACKGROUND

The most influential parts of an electric guitar are the strings, the magnetic pick-up, and the passive electrical tone control. Body resonances only have a minor influence on the resulting tone and will not be taken into account here. The guitar strings determine the basic sound since when vibrating, they are the primary sound source. The sound is mainly affected by the string material, tension, and stiffness. These features manifest primarily in frequency shifts of partial vibrations also known as the effect of inharmonicity [1]. The standard arrangement and tuning of a 6-string guitar with corresponding fundamental frequencies and MIDI number specifications is given in Table 1. Electromagnetic pick-ups

* All correspondance should be adressed to this author.

[1] http://www.songs2see.com/
[2] http://bandfuse.com/
[3] http://rocksmith.ubi.com/

capture the existing vibrations depending on their position on the instrument neck and the corresponding possible displacement of partials. Their technical specifications determine the transfer function which is commonly approximated by a second order low pass filter with a cut-off frequency in the range from 2 to 5 kHz. The same applies to the subsequent tone control of the guitar, which can be represented by a first order low pass filter. Both can be combined to an overall transfer function.

Table 1: *Standard Tuning of Guitar Strings.*

String Number	Standard Tuning	Fundamental Frequency	MIDI Number
1	E2	82.4 Hz	40
2	A2	110.0 Hz	45
3	D3	146.8 Hz	50
4	G3	196.0 Hz	55
5	B3	246.9 Hz	59
6	E4	329.6 Hz	64

Another important means of tone manipulation is the playing technique applied by the musician. In this work we distinguish 3 different plucking styles—finger style, picked, and muted—as well as and 5 expression styles—bending, slide, vibrato, harmonics, and dead notes—executed with the fingering hand in addition to non-decorated, normal expression style 2. See [2] for a detailed description of the playing techniques.

Table 2: *Playing Techniques.*

Plucking Style	Expression Style
finger style (F)	bending (BE)
picked (P)	slide (SL)
muted (M)	vibrato (VI)
	harmonics (HA)
	dead notes (DN)

Besides common music notation, a widespread method of notating guitar music is the tablature. By indicating the fret and string numbers to be used, it provides an alternative and more intuitive view of the played score. Figure 1 shows an example of a score and corresponding tablature.

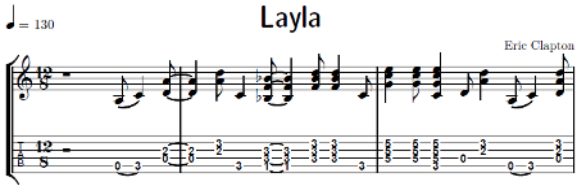

Figure 1: *Excerpt of the score and tablature representation of an interpretation from the Song Layla written by Eric Clapton [3].*

In tablature notation every drawn line symbolizes a string of the instrument, typically the lowest string corresponds to the bottom line. The numbers written on the single lines represent the used fret, where the fret number 0 corresponds to the open string.

3. PREVIOUS WORK

As will be discussed in Section 4, various Music Information Retrieval (MIR) tasks are relevant for our work. In the past, several authors focussed on *monophonic* guitar recordings, which contain isolated notes or simple melodies. The task of *onset detection*, i.e. the detection of note start times in audio recordings, was investigated in many publications. An overview over state-of-the-art methods can be found for instance in [4]. *Multipitch estimation*, i.e., the transcription of multiple simultaneously sounding notes, is up to this day a very challenging task to be performed in an automated manner [5]. In our paper, we build upon the method proposed by Fuentes et al. in [6]. For time-frequency-representation we use a spectral magnitude reassignment based on the instantaneous frequency as proposed in [7]. Fiss and Kwasinski proposed a multipitch estimation algorithm tailored towards the guitar in [8] by exploiting knowledge about the string tuning and pitch range of the instrument. Similarly, Yazawa et al. combine multipitch estimation with three constraints related to the guitar fretboard geometry to improve the transcription results [9]. In [3], an algorithm capable of real-time guitar string detection is presented, which is also the base for our work. Particularly for guitar chords, Barbancho et al. automatically classified between 330 different fingering configuration for three-voiced and four-voiced guitar chords by combining a multipitch estimation algorithm and a statistical modeling using a Hidden Markov Model (HMM) [10].

In addition to the score-based parametrization and the estimation of the fretboard position, we aim to estimate the *playing technique* that was used on the guitar to play each note. We showed in previous work, that the estimation of playing techniques [2] for electric bass guitar, which shares similar playing techniques with the electric guitar, can be performed from isolated note recordings with a high accuracy using a combination of audio features and machine learning techniques. Various publications analyzed guitar recordings with focus on playing techniques that modulate the fundamental frequency such as *vibrato* [11], *bending* [12], or *slides* [13, 12]. Other guitar playing techniques that were investigated in the literature are *slide*, *hammer-on*, and *pull-off* [13, 12]. A broader overview over state-of-the-art methods for the transcription and instrument-related parameters from string instrument recordings can be found in [14] and [15].

4. PROPOSED METHOD

4.1. Problem Formulation

The goal of this work is to develop an analysis algorithm, that extracts all essential parameters necessary for the automatic creation of guitar scores. Therefore, a robust event separation based on onset detection methods has to be implemented. Afterwards, the note duration and pitch must be extracted. In the next step, both the plucking and expression styles (see Table 2) as well as the string number must be estimated using feature extraction and subsequent classification methods. Finally, by using knowledge about the instrument string tuning, the fret position can be derived for each note.

The transcription parameters can be verified and corrected by exploiting knowledge about the instrument construction and physical limitations of the guitar player. Hence, a further goal is to develop adaptive algorithms that satisfy these conditions. The final model should be able to store the extracted parameters and

to generate a guitar tablature and score completely automatically based on a given polyphonic, monotimbral electric guitar recording. In this work exclusively clean guitar signals without any prior audio effect processing are considered. According to the diagram in Figure 2, the following sections will describe each step in detail.

4.2. Onset Detection

The purpose of this onset detection stage is the segmentation into musical note events. For the case of electric guitar recordings onsets corresponds to single plucks. The signal part between two plucks is interpreted as a note event. First, seven state-of-the-art onset detection functions (see appendix 8.1) were tested against a separate development set of guitar note recordings (see Section 5.1) using the same default blocksize and hopsize values. In general, these functions give an estimate of likelihood of a note onset to appear at each given time frame. Based on their superior performance, we selected the three best functions Spectral Flux, Pitchogram Novelty, and Rectified Complex Domain for the framework. Since all detection functions work in the frequency domain, we determined the optimal framesize for each function.

For the extraction of the onset positions, a peak picking algorithm proposed by Dixon [4] was used, which was optimized separately for each method. The results of each onset detection compared to the manually annotated ground truth are shown in Table 3. All detections are considered as true positives within an absolute tolerance area of 50 ms.

Table 3: *Optimal framesize and achieved F-measure for the best performing onset detection functions.*

Onset Detection Function	Optimal framesize	F-Measure
Spectral Flux	8 ms	0.93
Pitchogram Novelty	5 ms	0.87
Rectified Complex Domain	5 ms	0.95

The obtained onset positions of all detection functions are combined and filtered with an additional peak picking to avoid the detection of crackles, offsets, or duplicates that represent the same note onsets caused by this combination. Therefore, the mean square of energies $\bar{E}(n)$ in a variable interval τ before and after each onset candidate are analyzed and set into relation as

$$\bar{E}(n) = \frac{\sum_{i=-\tau}^{\tau} f(n+i)^2}{\tau}, \tag{1}$$

with $f(n)$ corresponding to the n^{th} frame of the summarized onset function f. With $\bar{E}_i$ denoting the mean squared energy of the ith interval ahead of the current onset, L corresponding to the length of the signal, f_s corresponding to the sampling frequency, and k_F and k_T being adjustment variables, the general conditions defining a detection as a valid onset are the following:

$$\min[\bar{E}_i(n-2\tau i)]_{i=1,2,3..I} < \bar{E}(n+\tau), \tag{2}$$

$$\frac{\sum_{i=1}^{L} f(i)^2}{L} < k_E \cdot \bar{E}(n+\tau) \tag{3}$$

and

$$n - n_{os(n-1)} > k_T \cdot f_s. \tag{4}$$

I is the maximum of intervals taken into account before the onset candidate, n is the sample index, and n_{os} is the sample number of the observed onset candidate.

In this work, the best results were achieved with $k_E = 100$, $k_T = 0.12$ ms, and $\tau = 331$ corresponding to 1.5 frames of 5 ms hopsize with a sample rate of 44100 Hz. The final method achieved an F-measure for onset detection of 98.5 %—all results are summarized in Section 5.3.

4.3. Multipitch Estimation

Next, the note segments of the audio signal are examined with respect to their spectral energy distribution. Large frame-sizes of $N = 4096$ and higher are necessary for the conventional Short-time Fourier Transform (STFT) to get a sufficient frequency resolution, which offers enough information for the pitch discrimination in the fundamental frequency register of the spectrum of a guitar. At the same time, large frame-sizes significantly reduce the achievable time resolution, which especially affects short notes. To avoid such complications, we compute a reassigned magnitude spectrogram based on the Instantaneous Frequency (IF) [7] representation in addition to the conventional time-frequency transform. By using the phase information for frequency correction, the IF supplies a high spectral accuracy while working with shorter frame sizes (here: $N = 1024$).

We use the IF magnitude spectrogram with a logarithmically-spaced frequency axis (84 bins per octave) as input for the subsequent Blind Harmonic Adaptive Decomposition (BHAD) algorithm proposed by Fuentes in [6]. It uses a frame overlap of 75 % and a downsampling by factor 4. The BHAD represents a multipitch estimation based on a framewise approach as previously used by Männchen et al. [3]. Several start frames (default: 5 frames) of each note event are left out to avoid the influence of noisy attack part transients. Furthermore, we aggregate over the following five frames in order to achieve more robust results. This way, note events with a minimum duration of 65 ms can be evaluated. For shorter events the amount of frames used for aggregation is reduced proportional.

We achieved an F-measure of 0.96 for pitch detection using this parameter setting. For the optimization of the algorithm concerning the number of aggregated frames and the parameters of the BHAD algorithm, we aimed at maximizing the Recall value (here: 0.98) in order to detect all possible fundamental frequency candidates. False positives are less critical since they can be eliminated by subsequent energy checks and checks of multiple pitch occurrences.

4.4. Partial Tracking

Based on the results of the pitch estimation, the fundamental frequency and the first 15 partials of each note event are tracked over time as follows. First, we apply a simple peak picking to the magnitude spectrum of each frame. The spectral peaks are assigned to harmonics of the different fundamental frequency candidates by minimizing the distance between the ideal harmonic frequency positions and the detected peak positions. We estimate the inharmonicity coefficient in each frame based on the detected partial peaks [3]. Results of the previous frames were used as initial inharmonicity values for the current frame and hence for more accurate partial estimation. The first frames were calculated with initial values based on [1].

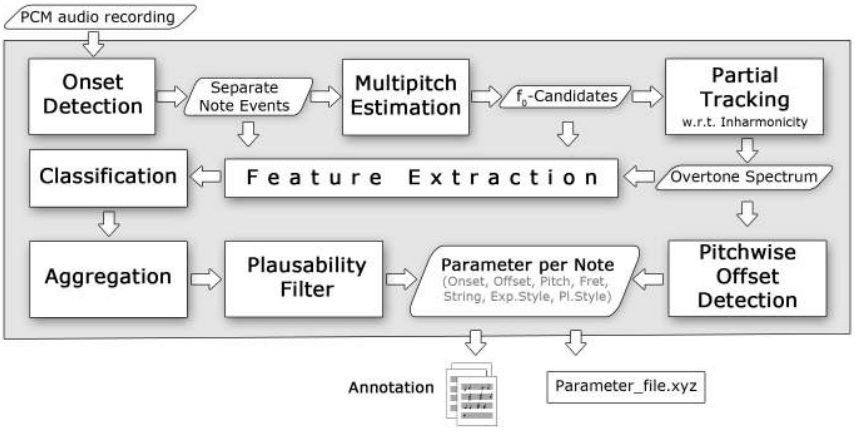

Figure 2: *Schematic model of the analysis framework.*

In addition, a window function is applied as a weighting function for variables of the tracking algorithm. These variables determine the frequency range around each partial where peaks are taken into account as well as the predicted frequency values of each partial for the following frame. A comparison of common window functions yields best performance for the use of a Kaiser function. The window length is fitted to the note duration and hence has the biggest impact in the middle of each note and almost no impact at the start and end position. Using this window, the considered search area for frequency peaks around each ideal harmonic frequency position is adjusted frame-wise. It adapts the extent of the range around each calculated partial which is taken into account for the performed peak picking to the relative note position.

Hence, at the temporal center of each note event this range is the largest and therefore more susceptible for frequency changes. Furthermore, the window function affects the amount of past frames taken into account when calculating the predicted harmonic frequencies of the current frame. At the center point of a note event less frames are considered emphasizing the affinity for frequency changes. Finally, the weight of magnitudes around each calculated harmonic frequency position is increased towards the middle of the note event. So, the comparison in the middle of note events yields lower dependency of the actual frequency distance but emphasizes high frequency magnitudes near the theoretical frequency. These three conditions are needed for an adaptive algorithm which reacts sensitive to frequency modulation techniques like bending, vibrato, and slide (see Section 2). A typical fundamental frequency envelope $f_0(t)$ for the frequency modulation technique *slide* is shown in Figure 3.

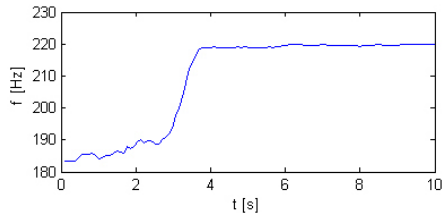

Figure 3: *Fundamental frequency envelope $f_0(t)$ of a slide tone.*

The obtained values allow for a correction of octave confusions, which can occur by the presence of weak sub-harmonics and their multiples, by summing up and comparing the odd and even harmonics of a note. Unlikely fundamentals are eliminated when the total energy of even partials falls below a quarter of the energy of odd partials.

4.5. Offset Detection

The detection of the note offset is performed based on the results of the partial tracking procedure as explained in the previous section. We obtain a temporal envelope for each time frame m by summing up the harmonic magnitude values $M_h(m)$ over all harmonics as

$$f_{\text{Env}}(m) = \sum_{h=1}^{H} M_h(m). \tag{5}$$

Figure 4 illustrates an example of a temporal magnitude envelope of a guitar note. The offset is obtained by detecting the first frame after the envelope peak with less than 5 % of the peak magnitude. Furthermore, an onset correction is performed by searching the lowest point of inflection before the peak. Therefore, the considered time area of the note excerpt is expanded in forward direction by 200 ms as safety margin. We initially smooth the envelope function by convolving it with a three-element-rectangle window to avoid the detection of random noise peaks.

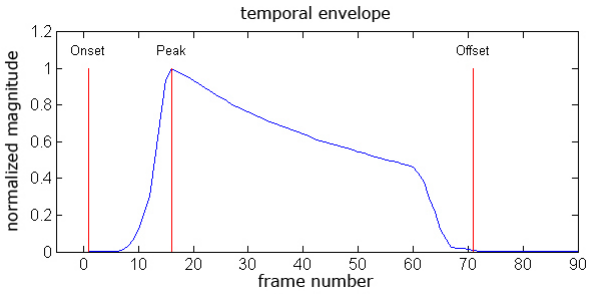

Figure 4: *Temporal magnitude envelope $f_{Env}(m)$ of the summed harmonics of a guitar note.*

4.6. Feature Extraction & Classification

Based on the extracted note parameters onset, offset, and pitch, various audio features can be extracted that allow to discriminate high-level parameters such as the played string or the applied playing techniques. We compute features on a frame-by-frame level and aggregate the features over the duration of each note event using different statistical measures such as minimum, maximum, mean, or median. A list of all 774 features can be found in appendix 8.2. A classification based on this amount of feature dimensions leads to high computational load and potential model overfitting. Therefore, prior to training the classification models, we first apply the feature selection algorithm *Inertia Ratio Maximization using Feature Space Projection* (IRMFSP) [16] for each classification task in order to reduce the dimensionality of the feature space. The amount of reduction depends on the performed classification task and is optimized separately.

For classification, we use a multi-class Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel. We perform three independent classification tasks—classification of the string number with 6 classes (see Table 1) as well as classification of the plucking style and expression style with three and six classes, respectively (see Table 2).

4.7. Plausability Filter

Depending on the classification task, the results can be aggregated over multiple notes using majority voting to obtain more robust classification. Furthermore, knowledge about typical guitar performance can be exploited in order to avoid impossible fingering positioning or improbable usage of playing techniques. In this section, two approaches to correct the classification results will be described. First, an *intra-note plausibility filter* deals with single notes and notes that are played simultaneously such as in chords. Second, an *inter-note plausibility filter* takes into account sequences of consecutive notes. Both filter aggregate classification results in dependence of the performed task for higher robustness. Expression styles are unified over all notes of a note event, plucking styles are aggregated over entire licks assuming that during one subsequently played lick no changes in plucking style occur.

4.7.1. Intra-Note Plausability Filter

The first corrections are applied to the estimated expression style. The most obvious restriction is a duration limit for the *dead note* class. Notes that are classified as dead notes and last longer than 0.6 seconds, are re-evaluated and the next probable expression style class is chosen. Second, we assume that all remaining expression styles (except the non-decorated *normal* expression style) are only plausible if less than three simultaneous pitches are detected by the multi-pitch estimation. Third, frequency modulated expression styles are compared against psychoacoustical thresholds so that detected bending, slide, or vibrato techniques with not noticeable frequency differences are set to the *normal* expression style. Especially for slide and bending, a check of the start and end pitch within the note is performed to detect and eliminate minor frequency changes below one semitone. Finally, the estimated pitches when marked as class *harmonics* are compared to possible harmonic pitches of the known guitar string tuning. Only certain pitches can occur at certain positions of the fretboard. Hence, *harmonics* detections with impossible pitch values are set to expression style *normal*.

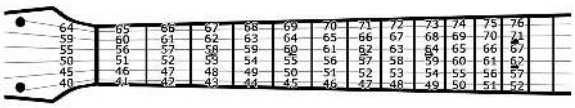

Figure 5: *Fretboard MIDI number references for standard tuning. The first column refers to empty string pitches.*

A second filter is applied to correct the string number. Each position on the fretboard is connected to a fixed pitch as shown in Figure 5, depending on the instrument tuning. Most pitch values within the pitch range of the guitar can be played on different fretboard position, hence, on different strings.

The first assumption of this filter connects to the expression style results by setting the probability of empty strings to zero if a decorating expression style has been classified. In addition, all probabilities from strings not allowing to play the observed pitch at any fretboard position are set to zero. Considering polyphonic fingerings, two or more pitches might collide by being assigned on the same string. To avoid this interception, our algorithm provides alternative fingering positions based on the string probabilities. Large spreads of the fingering are likely to occur as a result of alternative string assignment by using simple replacement to the most probable neighbour strings. Hence, spreads larger than four frets are eliminated. Instead, fingerings with smaller spreads are preferred by weighting their probabilities based on a computed fretboard centroid. Depending on the contribution of the fingering around the fretboard centroid the probability of each classification is lowered respectively to its relative fret and string distance to the centroid. Highest distances correspond to most intense lowering by half of the classification probability.

4.7.2. Inter-Note Plausability Filter

The inter-note plausability filter can correct classification results based on the parameters of preceding and subsequent notes. The first attempt of this filter is to find similar pitches played on the same string. Under the condition of comparable magnitudes and small gaps at the note borders notes are tied. As a consequence, detected expressions styles such as dead note become impossible for tied notes and are corrected. When comparing consecutive fingerings, fast and commonly applied position jumps (fingering changes with high local distances) are highly improbable if empty strings are not involved. Again, the fretboard centroid is used to weight and determine the most likely fingering if such jumps occur. This depends on the occurrence rate as well as the probability values of string estimation. The same corrections are performed for harmonic expression styles. Due to the characteristic of this playing technique, the fingering hand holds the string at an alternative fret position to obtain the perceived pitch. Here also the fretboard centroid is used to find the most probable position.

5. EVALUATION

For the evaluation of the proposed transcription algorithm, we use the common evaluation measures *Precision*, *Recall*, *F-Measure*, and *Accuracy*. In this section, a novel dataset of electric guitar recordings with extensive annotation of note parameters will be introduced. This dataset served as ground-truth in our experiments.

All results presented in Section 5.3 are based on 10-fold cross validation experiments.

5.1. Dataset

For the evaluation tasks, our novel dataset was recorded and manually annotated with all note parameters discussed in this paper. Six different guitars in standard tuning (see Table 1) were used with varying pick-up settings and different string measures to ensure a sufficient diversification in the field of electric guitars. The recording setup consisted of appropriate audio interfaces [4] which were directly connected to the guitar output. The recordings are provided in one channel RIFF WAVE format with 44100 Hz sample rate. The parameter annotations are stored in XML format.

The dataset consists of two sets. The first one created exclusively for this work contains all introduced playing techniques (see Table 2) and is provided with a bit depth of 24 Bit. It has been recorded using three different guitars and consists of about 4700 note events with monophonic and polyphonic structure. As a particularity the recorded files contain realistic guitar licks ranging from monophonic to polyphonic instrument tracks. In addition, a second set of data consisting of 400 monophonic and polyphonic note events with 3 different guitars is provided. No expression styles were applied here and each note event was recorded and stored in a separate file with a bit depth of 16 Bit [3]. The combined dataset will be made available as a public benchmark for guitar transcription research[5].

5.2. Experimental Procedure

For the onset detection, a detection within a tolerance of 50 ms to the annotated ground truth is considered as true positive. Since the offset detection is a harder task (due to smoothly decreasing note envelopes), a tolerance of 200 ms is used. Because of the time-frequency transform the duration of one additional frame (5 ms) has to be considered to obtain the effective tolerance of each temporal detection. The frequency tolerance adapts to the pitch and is scored as correct if both annotated and detected frequencies are rounded to the same MIDI pitch numbers. The three classification tasks discussed in Section 4.6 are measured using the mean normalized class accuracy.

5.3. Results

The performance of the final system for onset detection, offset detection, and pitch estimation are shown in Table 4. Because of the high specialization towards applications of guitar recordings the results clearly outperform existing approaches. Previous onset detection methods are on average placed around 90 % accuracy [4, 17], pitch estimation methods reached values up to 90 % [8, 5, 6].

The results of classification tasks are given in Table 5 - 7. In general, the typical decrease of accuracy for a higher number of classes can be observed. The string estimation still performed with good discrimination results of 82 % average accuracy including polyphonic estimation and the use of plausability filtering. The results differ from previous work [3, 10] where average accuracies around 90 % were reached due to different classification and evaluation methods. Plucking style estimation is performed with a very good score of 93 % average accuracy comparable to Abeßer et al. [2]. Here, a plausability filter was applied to combine the results of one note event. The classification of expression styles achieved good average accuracy of 83 %. State-of-the-art methods offer comparable results depending on the number of classes being distinguished. The plausability filter for expression styles introduced in Section 4.7 is used for correction and aggregation of the classification results.

[4] Tascam US 1641, M-Audio Fast Track Pro

[5] http://www.idmt.fraunhofer.de/en/business_units/smt/guitar.html

Table 4: *Precision, Recall and F-Measure results of onset detection, offset detection, and pitch estimation.*

Detection Function	Precision	Recall	F-Measure
Onset	0.98	0.99	0.99
Offset	0.98	0.98	0.98
Pitch Estimation	0.95	0.98	0.96

Table 5: *Accuracy results of the string estimation in percent displayed in a confusion matrix. Average accuracy = 82 %.*

string (correct) \ string (classified)	1	2	3	4	5	6
1	**81.3**	16.6	2.1	0.0	0.0	0.0
2	5.7	**86.0**	7.1	1.1	0.0	0.0
3	0.2	9.4	**78.8**	9.7	1.8	0.2
4	0.0	0.6	6.9	**81.8**	9.8	0.9
5	0.0	0.8	0.7	13.1	**76.7**	8.6
6	0.0	0.3	0.5	2.6	12.1	**84.5**

Table 6: *Accuracy results of the plucking style estimation in percent displayed in a confusion matrix. Average accuracy = 93 %. For abbreviations see Table 2.*

style (correct) \ style (classified)	F	P	M
F	**83.3**	16.7	0.0
P	2.5	**95.4**	2.0
M	1.9	1.9	**96.2**

With the automatically extracted transcription, guitar-specific tablature notation can be generated including information about the used playing techniques. A sample of the dataset is visualized in Figure 6. The tablature notation, which was automatically extracted from the audio recording, is compared against the reference notation taken from the dataset.

6. CONCLUSIONS

In this paper we introduced a novel algorithm for guitar transcription. The algorithm includes different estimation techniques for score-based and instrument-based parameters from isolated guitar recordings. By applying different optimization approaches, we received excellent detection results for onset, offset and pitch with an average accuracy of 96 % and higher. Estimations of more complex instrument-based parameters were performed with good results of 82 % and higher. Furthermore, a novel dataset was created and published to evaluate the proposed methods. We showed

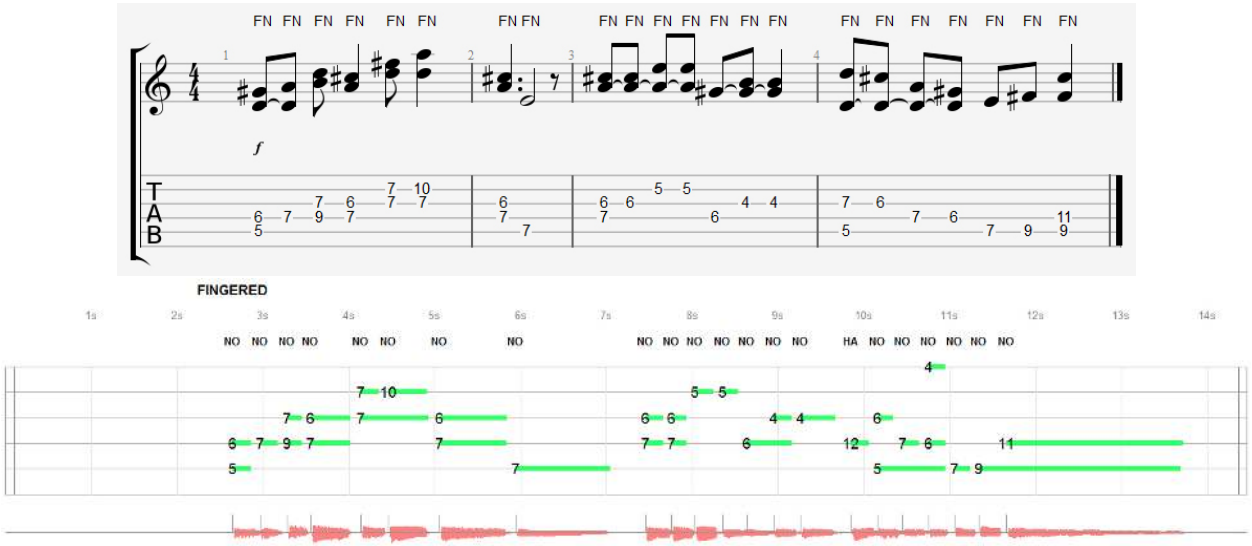

Figure 6: *Polyphonic guitar lick of the dataset.*
Top: *Manually notated tablature - Legend: FN above each note annotates the Plucking Style 'Finger Style' and Expression Style 'Normal'.*
Bottom: *Automatically notated tablature - Legend: Plucking Style is obtained for the entire lick. The letters above each note denote the Expression Style (NO - normal, HA - Harmonics).*

Table 7: *Accuracy results of the expression style estimation in percent displayed in a confusion matrix. Average accuracy = 83 %. For abbreviations see Table 2.*

stlye (correct) \ style (classified)	NO	BE	SL	VI	HA	DN
NO	**94.8**	0.7	0.5	0.9	1.5	1.6
BE	14.0	**71.3**	12.3	1.2	0.0	1.2
SL	20.7	11.2	**50.9**	8.6	4.3	4.3
VI	25.3	1.2	3.1	**66.7**	3.1	0.6
HA	10.5	0.0	0.0	2.0	**82.4**	5.2
DN	7.7	0.0	0.0	0.8	10.7	**80.8**

that an automatic transcription of guitar-based tablature is possible with a high accuracy.

7. REFERENCES

[1] Isabel Barbancho, Lorenzo J. Tardón, Simone Sammartino, and Ana M. Barbancho, "Inharmonicity-based method for the automatic generation of guitar tablature," in *Proceedings of the IEEE Transactions on Audio, Speech, and Language Processing*, August, 2012, pp. 1857–1868.

[2] Jakob Abeßer, Hanna Lukashevich, and Gerald Schuller, "Feature-based extraction of plucking and expression styles of the electric bass guitar," in *Proceedings of the Int. IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, USA, March 14-19, 2010, pp. 2290–2293.

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

[4] Simon Dixon, "Onset detection revisited," in *Proceedings of the 9th Int. Conference on Digital Audio Effects (DAFx)*, Montreal, Canada, September 18 - 20, 2006, pp. 133–137.

[5] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger Kirchhoff, and Anssi Klapuri, "Automatic music transcription: challenges and future directions.," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 407–434, 2013.

[6] Benoit Fuentes, Roland Badeau, and Gaël Richard, "Blind harmonic adaptive decomposition applied to supervised source separation," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, August 27 - 31, 2012, pp. 2654–2658.

[7] Toshihiko Abe, Takao Kobayashi, and Satoshi Imai, "Harmonics tracking and pitch extraction based on instantaneous frequency," in *Proceedings of the Int. IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Detroid, USA, May 9 - 12, 1995, pp. 756–759.

[8] Xander Fiss and Andres Kwasinski, "Automatic real-time electric guitar audio transcription," *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 373–376, 2011.

[9] Kazuki Yazawa, Daichi Sakaue, Kohei Nagira, Katsutoshi Itoyama, and Hiroshi G. Okuno, "Audio-based guitar tablature transcription using multipitch analysis and playability constraints," *Proceedings of the 38th IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 196–200, 2013.

[10] Ana M. Barbancho, Anssi Klapuri, Lorenzo J. Tardón, and Isabel Barbancho, "Automatic transcription of guitar chords

and fingering from audio," in *Proceedings of the IEEE Transactions on Speech and Language Processing 2012*, March, 2012, pp. 915–921.

[11] Cumhur Erkut, Matti Karjalainen, and Mikael Laurson, "Extraction of Physical and Expressive Parameters for Model-based Sound Synthesis of the Classical Guitar," in *Proceedings of the 108th Audio Engineering Society (AES) Convention*, 2000, pp. 19–22.

[12] Loïc Reboursière, Otso Lähdeoja, Thomas Drugman, Stéphane Dupont, Cécile Picard-Limpens, and Nicolas Riche, "Left and right-hand guitar playing techniques detection," in *Proceedings of the International Conference on New Interfaces for Muscial Expression (NIME)*, Ann Arbor, Michigan, USA, 2012, pp. 1–4.

[13] Tan Hakan Özaslan, Enric Guaus, Eric Palacios, and Josep Lluis Arcos, "Attack Based Articulation Analysis of Nylon String Guitar," in *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, Málaga, Spain, 2010, pp. 285–298.

[14] Jakob Abeßer, *Automatic Transcription of Bass Guitar Tracks applied for Music Genre Classification and Sound Synthesis*, Ph.D. thesis, Technische Universität Ilmenau, Germany, submitted: 2013.

[15] Christian Dittmar, Estefanía Cano, Sascha Grollmisch, Jakob Abeßer, Andreas Männchen, and Christian Kehling, "Music technology and music education," *Springer Handbook for Systematic Musicology*, 2014.

[16] Geoffroy Peeters and Xavier Rodet, "Hierarchical gaussian tree with inertia ratio maximization for the classification of large musical instrument databases," in *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx)*, London, UK, September 8-11, 2003, pp. 1–6.

[17] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler, "A tutorial on onset detection in music signals," in *Proceedings of the IEEE Transactions on Speech and Audio Processing*, September, 2005, pp. 1035–1047.

[18] Tan Hakan Özaslan and Josep Lluis Arcos, "Legato and Glissando Identification in Classical Guitar," in *Proceedings of Sound and Music Computing Conference (SMC), Barcelona, Spain*, 2010, pp. 457–463.

[19] Jakob Abeßer, "Automatic String Detection for Bass Guitar and Electric Guitar," *From Sounds to Music and Emotions - 9th International Symposium, CMMR 2012, London, UK, June 19-22, 2012, Revised Selected Papers*, pp. 333–352, 2013.

[20] Jakob Abeßer, Christian Dittmar, and Gerald Schuller, "Automatic recognition and parametrization of frequency modulation techniques in bass guitar recordings," in *Proceedings of the Audio Engeneering Society 42nd Int. Conference (AES)*, Ilmenau, Germany, July 22 - 24, 2011, pp. 1–8.

[21] Jakob Abeßer and Gerald Schuller, "Instrument-centered music transcription of bass guitar tracks," in *Proceedings of the Audio Engeneering Society 53rd Int. Conference (AES)*, London, UK, January 27 - 29, 2014, pp. 1–10.

[22] Loic Reboursière, Christian Frisson, Otso Lähdeoja, John Anderson Mills III, Cécile Picard, and Todor Todoroff, "Multimodal Guitar : A Toolbox For Augmented Guitar Performances," in *Proceedings of the Conference on New Interfaces for Musical Expression (NIME)*, Syndney, Australia, 2010, pp. 415–418.

[23] David Wagner and Stefan Ziegler, "Erweiterung eines Systems zur Detektion von Onsets in Musiksignalen," in *Media Project of the Technical University of Ilmenau*, Ilmenau, Germany, 2008.

[24] Christian Kehling, "Entwicklung eines parametrischen Instrumentencoders basierend auf Analyse und Re-Synthese von Gitarrenaufnahmen," Diploma thesis, Technical University of Ilmenau, Germany, 2013.

8. APPENDIX

8.1. List of Onset Detection Functions

In this work, we compared the onset detection functions Spectral Flux [4], Rectified Complex Domain [4], Weighted Phase Deviation [4], High Frequency Content [23], Modified Kullback-Leibler Distance [23], Foote [23], and Pitchogram Novelty [21].

8.2. Audio Features

Table 8: *Feature list for classification. If features generate more than one return-value the amount is written in brackets after the feature name. Novel features are marked bold.*

· Spectral Centroid	· Odd To Even Harmonic Energy Ratio
· **Relative Spectral Centroid**	· **Harmonic Spectral Centroid**
· Spectral Roll Off	· Harmonic Magnitude Slope
· Spectral Slope	· Relative Harmonic Magnitude (14)
· Spectral Spread	· Normalized Harmonics Frequency Deviation (14)
· Spectral Decrease	· Frequency Statistics: Maximum, Minimum, Mean, Median, Variance (5)
· Spectral Crest	· Frequency Statistics: Maximum, Minimum, Mean, Median, Variance (5)
· Spectral Flatness	
· Inharmonicity Factor	
· Tristimulus 1,2 und 3 (3)	
· Spectral Irregularity	

Each frame-based audio feature listed so far is condensed to 14 statistic values per note. Maximum, Minimum, Mean, Variance, Median, Skewness and Kurtosis are computed for the attack and the decay part of each note. Both are known durations from Section 4.5 because of the performed temporal refinement. In Addition several novel note-based features are appended to the feature vector:

· High Frequency Pre Onset Arousal	· Envelope Part Length (3)
· Magnitude Range	· Temporal Slope (2)
· Envelope Sum	· Range Attack Time Deviation
· Temporal Centroid	· Mean Attack Time Deviation
· Envelope Fluctuation(2)	· Variance Attack Time Deviation
· Envelope Modulation Frequency and Range	· Subharmonic Attack Energy (21)
	· Subharmonic Decay Energy (21)

Concatenation of all features yields a feature vector of 774 elements. The detailed computation steps are explained in [24].