Automatic String Detection for Bass Guitar and Electric Guitar

Jakob Abeßer

Fraunhofer IDMT,
Ehrenbergstr. 17, 98693 Ilmenau, Germany
{abr@idmt.fhg.de}
http://www.idmt.fraunhofer.de

Abstract. In this paper, we present a feature-based approach to automatically estimate the string number in recordings of the bass guitar and the electric guitar. We perform different experiments to evaluate the classification performance on isolated note recordings. First, we analyze how factors such as the instrument, the playing style, and the pick-up settings affect the performance of the classification system. Second, we investigate, how the classification performance can be improved by rejecting implausible classifications as well as aggregating the classification results over multiple adjacent time frames. The best results we obtained are f-measure values of $F = .93$ for the bass guitar (4 classes) and $F = .90$ for the electric guitar (6 classes).

Keywords: string classification, fretboard position, fingering, bass guitar, electric guitar, inharmonicity coefficient

1 Introduction

On string instruments such as the bass guitar or the guitar, most notes within the instrument's pitch range can be played at multiple positions on the instrument fretboard. Each *fretboard position* is defined by a unique string number and a fret number. Written music representations such as *common music notation* do not provide any information about the fretboard position where each note is to be played. Instead, musicians often have to choose an appropriate fretboard position based on their musical experience and stylistic preferences. The *tablature* representation, on the other hand, is specialized on the geometry of fretted string instruments such as the guitar or the bass guitar. It specifies the fretboard position for each note and thus resolves the ambiguity between note pitch and fretboard position. Fig. 1 illustrates a bass-line represented both as score and as tablature.

Conventional automatic music transcription algorithms extract score-related parameters such as the pitch, the onset, and the duration of each note. In order to analyze recordings of string instruments, the fretboard position needs to be estimated as an additional parameter. The ability to automatically estimate the fretboard position allows to generate a tablature and is therefore very useful for

9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)
19-22 June 2012, Queen Mary University of London

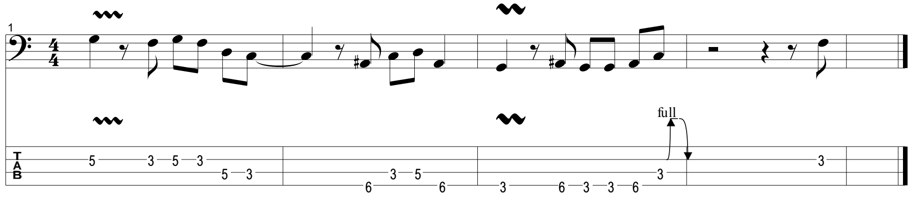


Fig. 1: Score and tablature representation of a bass-line. The four horizontal lines in the tablature correspond to the four strings with the tuning E1, A2, D2, and G2 (from bottom to top). The numbers correspond to the fret numbers on the strings that are to be played.

music assistance and music education software. This holds true especially if this software is used by beginners who are not familiar with reading musical scores. As will be discussed in Sect. 3, various methods for estimating the fretboard position were proposed in the literature so far, ranging from audio-based methods to methods that exploit the visual modality or that use attached sensors on the instrument. However, the exclusive focus on audio analysis methods for this purpose has several advantages: In music performance scenarios involving a bass guitar or electric guitar, the instrument signal is accessible since these instruments need to be amplified. In contrast, video recordings of performing musicians and the instrument neck are often limited in quality due to movement, shading, and varying lighting conditions on stage. Additional sensors that need to be attached to the instrument are often obtrusive to the musicians and affect their performance. Therefore, this paper focuses on a sole audio-based analysis.

This paper is structured as follows: We outline the goals and challenges of this work in Sect. 2. In Sect. 3, we discuss existing methods for estimating the fretboard position from string instrument recordings. A new approach solely based on audio-analysis is detailed in Sect. 4, starting with the spectral modeling of recorded bass and guitar notes in Sect. 4.1 and the note detection in Sect. 4.2. Based on the audio features explained in Sect. 4.2, we illustrate how the fretboard position is automatically estimated in Sect. 4.3. In Sect. 5, we present several evaluation experiments and discuss the obtained results. Finally, we conclude our work in Sect. 6.

2 Goals & Challenges

We aim to estimate the string number n_s from recorded notes of the bass guitar and the electric. Based on the note pitch P and the string number, we can apply knowledge on the instrument tuning to derive the fret number n_f and thus a complete description of the fretboard position. In the evaluation experiments described in Sect. 5, we investigate how the classification results are affected by separating the training and test data according to different criteria such as the instruments, the pick-up (PU) settings, and the applied playing techniques. Furthermore, we analyze if a majority voting scheme that combines multiple string classification results for each note can improve the classification performance. The main challenge is to identify suitable audio features that allow to discriminate between notes that, on the one hand, have the same fundamental frequency f_0 but, on the other hand, are played on different strings. The automatic classification of the played string is difficult since the change of fingering alters the sonic properties of the recorded music signal only subtly. Classic non-parametric spectral estimation techniques such as the Short-Time Fourier Transform (STFT) are affected by the *spectral leakage* effect: the Fourier Transform of the applied window function limits the achievable frequency resolution to resolve closely located spectral peaks. In order to achieve a sufficiently high frequency resolution for estimating the harmonic frequencies of a note, rather larger time frames are necessary. The decreased time resolution is disadvantageous if notes are played with frequency modulation techniques such as bending or vibrato, which cause short-term fluctuations of the harmonic frequencies [1]. This problem is especially impeding in lower frequency bands. Thus, a system based on classic spectral estimation techniques is limited to analyze notes with only a slow-varying pitch, which can be a severe limitation for a real-word system. Since we focus on the bass guitar and the electric guitar, frequencies between 41.2 Hz and 659.3 Hz need to be investigated as potential f_0-candidates[1].

3 Related Work

In this section, we discuss previous work on the estimation of the played string and the fretboard position from bass and guitar recordings. First, we review methods that solely focus on analyzing the audio signal. Special focus is put on the phenomenon of inharmonicity. Then, we compare different hybrid methods that incorporate computer vision techniques, instrument enhancements, and sensors.

3.1 Audio Analysis

Penttinen et al. estimated the plucking point on a string by analyzing the delay times of the two waves on the string, which travel in opposite directions after the string is plucked [21]. This approach solely focuses on a time-domain analysis and is limited towards monophonic signals. In [3], Barbancho et al. presented an algorithm to estimate the string number from isolated guitar note recordings. The instrument samples used for evaluation were recorded using different playing techniques, different dynamic levels, and guitars with different string material. After the signal envelope is detected in the time-domain, spectral analysis based on STFT is applied to extract the spectral peaks. Then, various audio features

[1] This corresponds to the most commonly used bass guitar string tunings E2 to G3 and electric guitar string tuning E3 to E5, respectively, and a fret range up to the 12th fret position.

related to the timbre of the notes are extracted such as the spectral centroid, the relative harmonic amplitudes of the first four harmonics, and the inharmonicity coefficient (compare Sect. 3.1). Furthermore, the temporal evolution of the partial amplitudes is captured by fitting an exponentially decaying envelope function. Consequently, only one feature vector can be extracted for each note. As will be shown in Sect. 4.2, the presented approach in this paper allows to extract one feature vectors on a frame-level. This allows to accumulate classification results from multiple (adjacent) frames of the same note recording to improve the classification performance (compare Sect. 4.3). The authors of [3] reported diverse results from the classification experiments. However, they did not provide an overall performance measure to compare against. The performance of the applied classification algorithm strongly varied for different note pitch values as well as for different compilations of the training set in their experiments.

In [2], Barbancho et al. presented a system for polyphonic transcription of guitar chords, which also allows to estimate the fingering of the chord on the guitar. The authors investigated 330 different fingering configuration for the most common three-voiced and four-voiced guitar chords. A Hidden Markov Model (HMM) is used to model all fingering configurations as individual hidden states. Based on an existing multi-pitch estimation algorithm, harmonic saliency values are computed for all possible pitch values within the pitch range of the guitar. Then, these saliency values are used as observations for the HMM. The transitions between different hidden states are furthermore constrained by two models—a musicological model, which captures the likelihood of different chord changes, and an acoustic model, which measures the physical difficulty of changing the chord fingerings. The authors emphasized that the presented algorithm is limited towards the analysis of solo guitar recordings. However, it clearly outperformed a state-of-the-art chord transcription system. The applied dataset contained instrument samples of electric guitar and acoustic guitar. Maezawa et al. proposed a system for automatic string detection from isolated bowed violin note recordings in [16]. Similar to the bass guitar, the violin has 4 different strings, but in a higher pitch range. The authors analyzed monophonic violin recordings of various classical pieces with given score information. First, the audio signal is temporally aligned to the musical score. For the string classification, filterbank energies are used as audio features and a Gaussian Mixture Model (GMM) as classifier. The authors proposed two additional steps to increase the robustness of the classification. First, feature averaging and feature normalization are used. Then, a context-dependent error correction is applied, which is based on empirically observed rules how musicians choose the string number. The authors investigated how training and test with the same and different instruments and string types affect the classification scores (similar to Sect. 5). The highest F-measure value that was achieved for the string classification with 4 classes is $F = .86$.

Inharmonicity For musical instruments such as the piano, the guitar, or the bass guitar, the equation describing the vibration of an ideal flexible string is extended by a restoring force caused by the string stiffness [7]. Due to dispersive wave propagation within the vibrating string, the effect of inharmonicity occurs, i.e., the purely harmonic frequency relationship of an ideal string is distorted and the harmonic frequencies are stretched towards higher values as

$$f_k = kf_0\sqrt{1+\beta k^2};\ k \geq 1 \tag{1}$$

with k being the harmonic index of each overtone and f_0 being the fundamental frequency. The inharmonicity coefficient β depends on different properties of the vibrating string such as Young's Modulus E, the radius of gyration K, the string tension T, the cross-sectional area S, as well as the string length L. With the string length being approximately constant for all strings of the bass guitar and the electric guitar, the string diameter usually varies from 0.45 mm to 1.05 mm for electric bass and from 0.1 mm to 0.41 mm for electric guitar[2]. The string tension T is proportional to the square of the fundamental frequency of the vibrating string. Järveläinen et al. performed different listening tests to investigate the audibility of inharmonicity towards humans [12]. They found that the human audibility threshold for inharmonicity increases with increasing fundamental frequency.

Hodgekinson et al. observed a systematic time-dependence of the inharmonicity coefficient if the string is plucked hard [10]. The authors found that β does not remain constant but increases over time for an acoustic guitar note. In contrast, for a piano note, no such behavior was observed. In this paper, we aim to estimate β on a frame-level and do not take the temporal evolution of β into account.

Different methods have been applied in the literature to extract the inharmonicity coefficient such as the cepstral analysis, the harmonic product spectrum [8], or inharmonic comb-filter [9]. For the purpose of sound synthesis, especially for physical modeling of string instruments, inharmonicity is often included into the synthesis models in order to achieve a more natural sound [24]. The inharmonicity coefficient of different instruments was analyzed as a distinctive feature in different Music Information Retrieval tasks such as instrument recognition and music transcription.

3.2 Hybrid Approaches & Visual Approaches

Different methods for estimating the fretboard position from guitar recordings were presented in the literature that include analysis methods from computer vision as a multi-modal extension of audio-based analysis.

A combined audio and video analysis was proposed by Hybryk and Kim to estimate the fretboard position of chords that were played on an acoustic guitar [11]. The goal of this paper was to first identify a played chord on the guitar regarding its "chord style", i.e., their root note and musical mode such as minor or major. For this purpose, the Specmurt [22] algorithm was used for

[2] These values correspond to commonly used string gauges.

spectral analysis in order to estimate a set of fundamental frequency candidates that can be associated to different note pitches. Based on the computed "chord style" (e.g., E minor), the "chord voicing" was estimated by tracking the spatial position of the hand on the instrument neck. The chord voicing is similar to the chord fingering as described in [2].

Another multi-modal approach for transcribing acoustic guitar performances was presented by Paleari et al. in [19]. In addition to audio analysis, the visual modality was analyzed to track the hand of the guitar player during his performance to estimate the fretboard position. The performing musicians were recorded using both two microphones and a digital video camera. The fretboard was first detected and then spatially tracked over time.

Other approaches solely used computer vision techniques for spatial transcription. Burns and Wanderley presented an algorithm for real-time finger-tracking in [4]. They used *attached cameras* on the guitar in order to get video recordings of the playing hand on the instrument neck. Kerdvibulvech and Saito used a stereo-camera setup to record a guitar player in [13]. Their system for finger-tracking requires the musician to wear *colored fingertips*. The main disadvantage of all these approaches is that both the attached cameras as well as the colored fingertips are unnatural for the guitar player. Therefore, they likely limit and impede the musician's expressive gestures and playing style.

Enhanced music instruments are equipped with additional sensors and controllers in order to directly measure the desired parameters instead of estimating them from the audio or video signal. On the one hand, these approaches lead to a high detection accuracy. On the other hand, these instrument extensions are obtrusive to the musicians and can affect their performance on the instrument [11]. In contrast to regular electric guitar pickups, *hexaphonic pickups* separately capture each vibrating string. In this way, spectral overlap between the string signals is avoided, which allows a fast and robust pitch detection with very low latency and very high accuracy, as shown for instance by O'Grady and Rickard in [18].

4 New Approach

Fig. 2 provides an overview over the string classification algorithm proposed in this paper. All processing steps are explained in detail in the next sections.

4.1 Spectral Modeling

Non-parametric spectral estimation methods such as the Periodogram make no explicit assumption on the type of signal that is analyzed. In order to obtain a sufficiently high frequency resolutions for a precise f_0-detection, relatively large time frames of data samples are necessary in order to compensate the spectral leakage effect, which is introduced by windowing the signal into frames. In contrast to the percussive nature of its short attack part (between approx. 20 ms and 40 ms), the decay part of a plucked string note can be modeled by a sum of decaying sinusoidal components. Their frequencies have a nearly perfectly

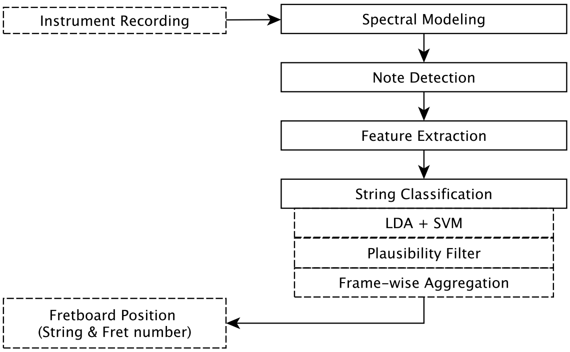


Fig. 2: Algorithm overview

harmonic relationship. Since the strings of the bass guitar and the electric guitar have a certain amount of stiffness, the known phenomenon of inharmonicity appears (compare Sect. 3.1).

Parametric spectral estimation techniques can be applied if the analyzed signal can be assumed to be generated by a known model. In our case, the power spectral density (PSD) $\Phi(\omega)$ can be modeled by an auto-regressive (AR) filter such as

$$\Phi(\omega) \approx \Phi_{AR}(\omega) = \sigma^2 \left| \frac{1}{1 + \sum_{l=1}^{p} a_l e^{-jl\omega}} \right|^2 \tag{2}$$

with σ^2 denoting the process variance, p denoting the model order, and $\{a_l\} \in \mathbb{R}^{p+1}$ being the filter coefficients. Since auto-regressive processes are closely related to linear prediction (LP), both a *forward prediction error* and a *backward prediction error* can be defined to measure the predictive quality of the AR filter. We use the *least-squares method* (also known as *modified covariance method*) for spectral estimation. It is based on a simultaneous least-squares minimization of both prediction errors with respect to all filter coefficients $\{a_l\}$. This method has been shown to outperform related algorithms such as the Yule-Walker method, the Burg algorithm, and the covariance method (See [17] for more details). The size of the time frames N is only restricted by the model order as $p \leq 2N/3$.

First, we down-sample the signals to $f_s = 5.5$ kHz for the bass guitar samples and $f_s = 10.1$ kHz for the electric guitar samples. This way, we can detect the first 15 harmonics of each note within the instrument pitch ranges, which is necessary for the subsequent feature extraction as explained in Sect. 4.2. In Fig. 3, the estimated AR power spectral density for a bass guitar sample (E1) as well as the estimated partials are illustrated. Since we only focus on isolated instrument samples here, we assume the fundamental frequency f_0 to be known in advance. The separate evaluation of fundamental frequency estimation is not within the scope of this paper.

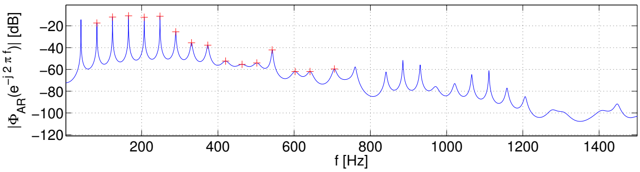


Fig. 3: Estimated AR power spectral density for the bass guitar sample with pitch E1 ($f_0 = 44.1Hz$). The estimated first 15 partials are indicated with red crosses.

By using overlapping time frames with a block-size of $N = 256$ and a hop-size of $H = 64$, we apply the spectral estimation algorithm to compute frame-wise estimates of the filter coefficients $\{a_l(n)\}$ in the frames that are selected for analysis (compare Sect. 4.2). In order to estimate the harmonic frequencies $\{f_k\}$, we first compute the pole frequencies of the AR filter by computing the roots of the numerator in Eqn. (2). Then, we assign one pole frequency to each harmonic according to the highest proximity to its theoretical frequency value as computed using Eqn. (1).

4.2 Feature Extraction

Note Detection In Sect. 4.1, we discussed that notes played on the bass guitar and the guitar follow a signal model of decaying sinosoidal components, i.e., the partials. In this section, we discuss how we extract audio features that capture the amplitude and frequency characteristics. We first detect the first frame shortly after the note attack part of the note is finished and the harmonic decay part begins. As mentioned in Sect. 4.1, signal frames with a percussive characteristic are indicated by high values of the process variance $\sigma^2(t)$ obtained the AR spectral estimation. We found that time frames after

$$t^\star = \arg\max_t \sigma^2(t) \tag{3}$$

are suitable for feature extraction. If the aggregation of multiple frame-wise results is used, we extract features in the first 5 frames after $t^\star$.

Inharmonicity estimation In each analyzed frame, we estimate the discrete frequencies f_k of the first 15 partials. Then, we estimate the inharmonicity coefficient β_k as follows. From Eq. (1), we obtain

$$(f_k/f_0)^2 = k^2 + \beta k^4 \tag{4}$$

We use polynomial curve fitting to approximate the left-hand side of Eq. (4) by a polynomial function of order 4 as

$$(f_k/f_0)^2 \approx \sum_{i=0}^{4} p_i k^i \tag{5}$$

Feature	Feature dimension
Inharmonicity coefficient $\hat{\beta}$	1
Relative partial amplitudes $\{\hat{a}_{r,k}\}$	15
Statistics over $\{\hat{a}_{r,k}\}$	8
Normalized partial frequency deviations $\{\Delta\hat{f}_{norm,k}\}$	15
Statistics over $\{\Delta\hat{f}_{norm,k}\}$	8
Partial amplitude slope $\hat{s}_a$	1
All features	$\sum = 48$

Table 1: Overview of all applied audio features.

and use the coefficient p_4 as an estimate of the inharmonicity coefficient β:

$$\hat{\beta} \approx p_4 \tag{6}$$

Partial-based Features In addition to the inharmonicity coefficient β, we compute various audio features that capture the amplitude and frequency characteristics of the first 15 partials of a note. First, we compute the relative amplitudes

$$\{\hat{a}_{r,k}\} = \{a_k/a_0\} \tag{7}$$

of the first 15 partials related to the amplitude of the fundamental frequency. Then, we approximate the relative partial amplitude values $\{\hat{a}_{r,k}\}$ as a linear function over k as

$$\hat{a}_{r,k} \approx p_1 k + p_0 \tag{8}$$

by using linear regression. We use the feature $\hat{s}_a = p_1$ as estimate of the *spectral slope* towards higher partial frequencies.

Based on the estimated inharmonicity coefficient $\hat{\beta}$ and the fundamental frequency f_0, we compute the theoretical partial frequency values $\{f_{k,theo}\}$ of the first 15 partials based on Eq. (1) as

$$f_{k,theo} = k f_0 \sqrt{1 + \hat{\beta} k^2}. \tag{9}$$

Then, we compute the deviation between the theoretical and estimated partial frequency values and normalize this difference value as

$$\Delta\hat{f}_{norm,k} = \frac{f_{k,theo} - \hat{f}_k}{\hat{f}_k}. \tag{10}$$

Again, we compute $\{\Delta\hat{f}_{norm,k}\}$ for the first 15 partials and use them as features. In addition, we compute the statistical descriptors maximum value, minimum value, mean, median, mode (most frequent sample), variance, skewness, and kurtosis over both $\{\hat{a}_{r,k}\}$ and $\{\Delta\hat{f}_{norm,k}\}$. Tab. 1 provides an overview over all dimensions of the feature vectors.

4.3 Estimation Of The Fretboard Position

String Classification In order to automatically estimate the fretboard position from a note recording, we first aim to estimate the string number n_s. Therefore, we compute the 48-dimensional feature vector $\{x_i\}$ as described in the previous section. We use Linear Discriminant Analysis (LDA) to reduce the dimensionality of the feature space to $N_d = 3$ dimensions for bass guitar and to $N_d = 5$ dimensions for guitar[3]. Then we train a Support Vector Machine (SVM) classifier using a Radial Basis Function (RBF) kernel with the classes defined by notes played on each string. SVM is a binary discriminative classifier that attempts to find an optimal decision plane between feature vectors of the different training classes [25]. The two kernel parameters C and γ are optimized based on a three-fold grid search. We use the LIBSVM library for our experiments [5].

The SVM returns probabilities $\{p_i\}$ to assign unknown samples to each string class. We estimate the string number $\hat{n}_s$ as

$$\hat{n}_s = \arg\max_i \{p_i\}. \tag{11}$$

We derive the the fret number $\hat{n}_f$ from the estimated string number $\hat{n}_s$ by using knowledge on the instrument tuning as follows. The common tuning of the bass is E1, A2, D2, and G2; the tuning of the guitar is E2, A2, D3, G3, B3, and E3. The string tunings can be directly translated into a vector of corresponding MIDI pitch values as $\{P_T\} = [28, 33, 38, 43]$ and $\{P_T\} = [40, 45, 50, 55, 59, 64]$, respectively.

In order to derive the fret number $\hat{n}_s$, we first obtain the MIDI pitch value P that corresponds to the fundamental frequency f_0 as

$$P = \lfloor 12 \log_2(f_0/440) - 69 \rfloor \tag{12}$$

Given the estimated string number $\hat{n}_s$, the fret number can be computed as

$$\hat{n}_f = P - P_T(\hat{n}_s). \tag{13}$$

A fret number of $\hat{n}_f = 0$ indicates that a note was played by plucking an open string.

Plausibility Filter As mentioned earlier, most note pitches within the frequency range of both the bass guitar and the guitar can be played on either one, two, or three different fret positions on the instrument neck. The pitch ranges are E2 to G3 for the bass guitar and E3 to E5 for the electric guitar considering a fret range up to the 12th fret position. Based on knowledge about the instrument tunings, we can derive a set of MIDI pitch values that can be played on each string. Therefore, for each estimated MIDI pitch value $\hat{P}$, we can derive a list of strings, where this note can theoretically be played on. If the plausibility filter is used, we set the probability values in $\{p_i\}$ of all strings, where this note can not be played on to 0 before estimating the string number as shown in Eq. (11).

[3] The number of dimensions N_d is chosen as $N_d = N_{strings} - 1 \equiv N_{classes} - 1$.

Aggregation of multiple classification results If the result aggregation is used, we sum up all class probability values $\{p_i\}$ over 5 adjacent frames. Then we estimate the string number as shown in Eq. (11) over the accumulated probability values.

5 Evaluation & Results

5.1 Dataset

For the evaluation experiments, we use a dataset of 1034 audio samples. These samples are isolated note recordings, which were taken from the dataset previously published in [23].[4] The samples were recorded using two different bass guitars and two different electric guitars, each played with two different plucking styles (plucked with a plectrum and plucked with the fingers) and recorded with two different pick-up settings (either neck pick-up or body pick-up).

5.2 Experiments & Results

Experiment 1: Feature Selection for String Classification In this experiment, we aim to identify the most discriminant features for the automatic string classification task as discussed in Sect. 4.3. Therefore, we apply the feature selection algorithm Inertia Ratio Maximization using Feature Space Projection (IRMFSP) [15, 20] to all feature vectors and the corresponding class labels separately for both instrument. In Tab. 2, the five features that are first selected by the IRMFSP algorithm are listed for the bass guitar and the electric guitar.

The features $\Delta\hat{f}_{norm}$, $\hat{\beta}$, and $\hat{a}_{r,k}$ as well as the derived statistic measures were selected consistently for both instruments. These features measure frequency and amplitude characteristics of the partials and show high discriminative power between notes played on different strings independently of the applied instrument. The boxplots of the two most discriminative features $\Delta f_{norm,9}$ for bass and $\Delta f_{norm,15}$ for guitar are illustrated separately for each instrument string in Fig. 4.

Since the deviation of the estimated harmonic frequencies from their theoretical values apparently carries distinctive information to discern between notes on different instrument strings, future work should investigate, if Eq. (1) could be extended by higher order polynomial terms in order to better fit to the estimated harmonic frequency values.

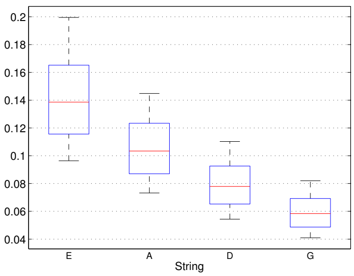
Experiment 2: String Classification in different conditions In this experiment, we aim to investigate how the performance of the automatic string classification algorithm is affected by

[4] This dataset contains isolated notes from bass guitar and electric guitar processed with various audio effects. In this work, only the non-processed note recordings were used.

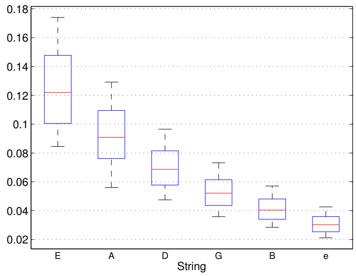
Rank	Bass Guitar	Electric Guitar
1	$\Delta\hat{f}_{norm,9}$	$\Delta\hat{f}_{norm,15}$
2	$\hat{\beta}$	mean$\{\hat{a}_{r,k}\}$
3	$\Delta\hat{f}_{norm,3}$	var$\{\Delta\hat{f}_{norm,k}\}$
4	var$\{\Delta\hat{f}_{norm,k}\}$	max$\{\hat{a}_{r,k}\}$
5	$\hat{a}_{r,4}$	skew$\{\Delta\hat{f}_{norm,k}\}$

Table 2: Most discriminative audio features for the string classification task as discussed in Sect. 5.2. Features are given in order as selected by the IRMFSP algorithm.

- the separation of the training and test set according to the applied instrument, playing technique, and pick-up setting,
- the instrument / the number of string classes,
- the use of a plausibility filter (compare Sect. 4.3),
- and the use of a aggregation of multiple classification results for each sample (compare Sect. 4.3).



(a) Boxplot of feature $\Delta f_{norm,9}$ for bass.



(b) Boxplot of feature $\Delta f_{norm,15}$ for guitar.

Fig. 4: Boxplots of the two most discriminative features for bass guitar and electric guitar.

The different conditions are illustrated in Tab. 3 for the bass guitar and in Tab. 4 for the electric guitar. The colums "Separated instruments", "Separated playing techniques", and "Separated pick-up setting" indicate which criteria were applied to separate the samples from training and test set in each configuration. The fifth and sixth column indicate whether the plausibility filter and the frame result aggregation were applied. In the seventh column, the number of folds for the configuration 1.6 and 2.6 and the number of permutations for the remaining configurations are given. The evaluation measures precision, recall, and F-measure were always averaged over all permutations and all folds, respectively.

Experiment	Separated instruments	Separated playing techniques	Separated pick-up settings	Plausibility filter (see Sect. 4.3)	Result aggregation over 5 frames (see Sect. 4.3)	No. of Permutations$^\diamond$/ No. of CV folds*	Precision $\bar{P}$	Recall $\bar{R}$	F-Measure $\bar{F}$
1.1.a	x					$2^\diamond$	.85	.85	.85
1.1.b	x			x		$2^\diamond$	.87	.87	.87
1.1.c	x			x	x	$2^\diamond$	.78	.78	.78
1.2.a	x	x				$8^\diamond$	.86	.86	.86
1.2.b	x	x		x		$8^\diamond$	.87	.87	.87
1.2.c	x	x		x	x	$8^\diamond$	.88	.88	.88
1.3.a		x	x			$8^\diamond$	.57	.50	.49
1.3.b		x	x	x		$8^\diamond$	.71	.69	.69
1.3.c		x	x	x	x	$8^\diamond$	.88	.88	.88
1.4.a		x				$8^\diamond$	.60	.54	.54
1.4.b		x		x		$8^\diamond$	.73	.71	.72
1.4.c		x		x	x	$8^\diamond$	**.93**	**.93**	**.93**
1.5.a			x			$8^\diamond$	.62	.55	.54
1.5.b			x	x		$8^\diamond$	.74	.71	.71
1.5.c			x	x	x	$8^\diamond$	.92	.92	.92
1.6.a						$10^\star$	.92	.92	.92
1.6.b				x		$10^\star$	.93	.93	.93
1.6.c				x	x	$10^\star$	**.93**	**.93**	**.93**

Table 3: Mean Precision $\bar{P}$, mean Recall $\bar{R}$, and mean F-Measure $\bar{F}$ for different evaluation conditions (compare Sect. 5.2) for the bass guitar.

After the training set and the test set are separated, the columns of the training feature matrix were first normalized to zero mean and unit variance. The mean vector and the variance vector were kept for later normalization of the test data. Subsequently, the normalized training feature matrix is used to derive the transformation matrix via LDA. We chose $N = N_{Strings} - 1$ as number of feature dimensions. The SVM model is then trained using the projected training feature matrix and a two-dimensional grid search is performed to determine the optimal parameters C and γ as explained in Sect. 4.3. For the configurations 1.6 and 2.6, none of the criteria to separate the training and the test set was applied. Instead, here we used a 10-fold cross-validation and averaged the precision, recall, and F-measure over all folds.

The results shown in Tab. 3 and Tab. 4 clearly show that both the plausibility filter as well as the result aggregation step significantly improve the classification results in most of the investigated configurations. Furthermore, we can see that the separation of training and test samples according to instrument, playing technique, and pick-up setting has a strong influence on the achievable classification performance. In general, the results obtained for the bass guitar and the electric guitar show the same trends. We obtain the highest classification

Experiment	Separated instruments	Separated playing techniques	Separated pick-up settings	Plausibility filter (see Sect. 4.3)	Result aggregation over 5 frames (see Sect. 4.3)	No. of Permutations$^\diamond$/ No. of CV folds*	Precision $\bar{P}$	Recall $\bar{R}$	F-Measure $\bar{F}$
2.1.a	x					$2^\diamond$	.64	.64	.63
2.1.b	x			x		$2^\diamond$	.70	.70	.70
2.1.c	x			x	x	$2^\diamond$	.76	.75	.75
2.2.a	x	x				$8^\diamond$	.69	.69	.68
2.2.b	x	x		x		$8^\diamond$	.71	.71	.70
2.2.c	x	x		x	x	$8^\diamond$	.78	.77	.77
2.3.a		x	x			$8^\diamond$	.61	.57	.56
2.3.b		x	x	x		$8^\diamond$	.68	.66	.66
2.3.c		x	x	x	x	$8^\diamond$	.74	.74	.73
2.4.a		x				$8^\diamond$	.64	.61	.60
2.4.b		x		x		$8^\diamond$	.71	.69	.69
2.4.c		x		x	x	$8^\diamond$	.80	.79	.79
2.5.a			x			$8^\diamond$	.69	.65	.65
2.5.b			x	x		$8^\diamond$	.74	.72	.72
2.5.c			x	x	x	$8^\diamond$	.84	.84	.84
2.6.a						$10^\star$	.72	.69	.70
2.6.b				x		$10^\star$	.81	.81	.81
2.6.c				x	x	$10^\star$	**.90**	**.90**	**.90**

Table 4: Mean Precision $\bar{P}$, mean Recall $\bar{R}$, and mean F-Measure $\bar{F}$ for different evaluation conditions (compare Sect. 5.2) for the electric guitar.

scores—$\bar{F} = .93$ for the bass guitar (4 classes) and $\bar{F} = .90$ for the electric guitar (6 classes)—for the configurations 1.6 and 2.6. These results indicate that the presented method can be successfully applied in different application tasks that require an automatic estimation of the played instrument string. In contrast to [16], we did not make use any knowledge about the musical context such as derived from a musical score.

We performed a baseline experiment separately for both instruments using Mel Frequency Cepstral Coefficients (MFCC) as features as well as LDA and SVM for feature space transformation and classification, respectively (compare Sect. 4.3). The same experimental conditions as in configuration 1.6. and 2.6. (see Sect. 5.2) were used. The classification results were performed and evaluated on a frame level. A 10-fold stratified cross-validation was applied and the results were averaged over all folds. We achieved classification scores of $\bar{F} = .46$ for the bass guitar and $\bar{F} = .37$ for electric guitar.

6 Conclusions

In this paper, we performed several experiments towards the automatic classification of the string number from given isolated note recordings. We presented a selection of audio features that can be extracted on a frame-level. In order to improve the classification results, we first apply a plausibility filter to avoid non-meaningful classification results. Then, we use an aggregation of multiple classification results that are obtained from adjacent frames of the same note. Highest string classification scores of $\bar{F} = .93$ for the bass guitar (4 string classes) and $\bar{F} = .90$ for the electric guitar (6 string classes) were achieved. As shown in a baseline experiment, classification systems based on commonly-used audio features such as MFCC were clearly outperformed for the given task.

7 Acknowledgements

The author likes to thank Michael Stein for the use of his data set. The Thuringian Ministry of Economy, Employment and Technology supported this research by granting funds of the European Fund for Regional Development to the project *Songs2See*[5], enabling transnational cooperation between Thuringian companies and their partners from other European regions.

References

1. J. Abeßer, C. Dittmar, and G. Schuller. Automatic Recognition and Parametrization of Frequency Modulation Techniques in Bass Guitar Recordings. In *Proc. of the 42nd Audio Engineering Sociery (AES) International Conference on Semantic Audio*, pages 1–8, Ilmenau, Germany, 2011.
2. A. M. Barbancho, A. Klapuri, L. J. Tardón, and I. Barbancho. Automatic Transcription of Guitar Chords and Fingering from Audio. *IEEE Transactions on Audio, Speech, and Language Processing*, pages 1–19, 2011.
3. I. Barbancho, A. M. Barbancho, L. J. Tardón, and S. Sammartino. Pitch and Played String Estimation in Classic and Acoustic Guitars. In *Proceedings of the 126th Audio Engineering Society (AES) Convention, Munich, Germany*, 2009.
4. A. Burns and M. Wanderley. Visual Methods for the Retrieval of Guitarist Fingering. In *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06)*, pages 196–199, Paris, France, 2006.
5. C.-C. Chang and C.-J. Lin. LIBSVM : A Library for Support Vector Machines. Technical report, Department of Computer Science, National Taiwan University, Taipei, Taiwan, 2011.
6. M. G. Christensen and A. Jakobsson. *Multi-Pitch Estimation.* Synthesis Lectures on Speech and Audio Processing. Morgan & Claypool Publishers, 2009.
7. N. H. Fletcher and T. D. Rossing. *The Physics Of Musical Instruments.* Springer, New York, London, 2 edition, 1998.
8. A. Galembo and A. Askenfelt. Measuring inharmonicity through pitch extraction. *Speech Transmission Laboratory. Quarterly Progress and Status Reports (STL-QPSR)*, 35(1):135–144, 1994.

[5] `http://www.songs2see.net`

9. A. Galembo and A. Askenfelt. Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano. *IEEE Transactions on Speech and Audio Processing*, 7(2):197–203, 1999.
10. M. Hodgkinson, J. Timoney, and V. Lazzarini. A Model of Partial Tracks for Tension-Modulated Steel-String Guitar Tones. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFX-10), Graz, Austria*, number 1, pages 1–8, 2010.
11. A. Hrybyk and Y. Kim. Combined Audio and Video for Guitar Chord Identification. In *Proc. of the 11th International Society for Music Information Retrieval Conference (ISMIR), Utrecht, Netherlands*, number Ismir, pages 159–164, 2010.
12. H. Järveläinen, V. Välimäki, and M. Karjalainen. Audibility of the timbral effects of inharmonicity in stringed instrument tones. *Acoustics Research Letters Online*, 2(3):79, 2001.
13. C. Kerdvibulvech and H. Saito. Vision-Based Guitarist Fingering Tracking Using a Bayesian Classifier and Particle Filters. *Advances in Image and Video Technology*, pages 625–638, 2007.
14. A. Klapuri. Multipitch analysis of polyphonic music and speech signals using an auditory model. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):255–266, 2008.
15. H. Lukashevich. Feature selection vs. feature space transformation in automatic music genre classification tasks. In *Proc. of the AES Convention*, 2009.
16. A. Maezawa, K. Itoyama, T. Takahashi, T. Ogata, and H. G. Okuno. Bowed String Sequence Estimation of a Violin Based on Adaptive Audio Signal Classification and Context-Dependent Error Correction. In *Proc. of the 11th IEEE International Symposium on Multimedia (ISM2009)*, pages 9–16, 2009.
17. S. L. Marple. *Digital Spectral Analysis With Applications.* Prentice Hall, Australia, Sydney, 1987.
18. P. D. O'Grady and S. T. Rickard. Automatic Hexaphonic Guitar Transcription Using Non-Negative Constraints. In *Proc. of the IET Irish Signals and Systems Conference (ISSC)*, pages 1–6, Dublin, Ireland, 2009.
19. M. Paleari, B. Huet, A. Schutz, and D. Slock. A Multimodal Approach to Music Transcription. In *Proc. of the 15th IEEE International Conference on Image Processing (ICIP)*, pages 93–96, 2008.
20. G. Peeters and X. Rodet. Hierarchical gaussian tree with inertia ratio maximization for the classification of large musical instruments databases. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx)*, London, UK, 2003.
21. H. Penttinen and J. Siiskonen. Acoustic Guitar Plucking Point Estimation in Real Time. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 209–212, 2005.
22. S. Saito, H. Kameoka, K. Takahashi, T. Nishimoto, and S. Sagayama. Specmurt Analysis of Polyphonic Music Signals. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(3):639–650, Feb. 2008.
23. M. Stein, J. Abeßer, C. Dittmar, and G. Schuller. Automatic Detection of Audio Effects in Guitar and Bass Recordings. In *Proceedings of the 128th Audio Engineering Society (AES) Convention, London, UK*, 2000.
24. V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen. Discrete-time modelling of musical instruments. *Reports on Progress in Physics*, 69(1):1–78, Jan. 2006.
25. V. N. Vapnik. *Statistical learning theory.* Wiley New York, 1998.