FEATURE-BASED EXTRACTION OF PLUCKING AND EXPRESSION STYLES OF THE ELECTRIC BASS GUITAR

Jakob Abeßer, Hanna Lukashevich, Gerald Schuller

Fraunhofer IDMT, Ilmenau, Germany. E-mail: {abr, lkh, shl}@idmt.fraunhofer.de

ABSTRACT

In this paper, we present a feature-based approach for the classification of different playing techniques in bass guitar recordings. The applied audio features are chosen to capture typical instrument sounds induced by 10 different playing techniques. A novel database that consists of approx. 4300 isolated bass notes was assembled for the purpose of evaluation. The usage of domain-specific features in a combination of feature selection and feature space transformation techniques improved the classification accuracy by over 27% points in comparison to a state-of-the-art baseline system. Classification accuracy reached 93.25% and 95.61% for the recognition of plucking and expression styles respectively.

Index Terms—electric bass guitar, transcription, plucking style, expression style, expressive performance analysis

1. INTRODUCTION

Common automatic music transcription algorithms focus on the extraction of the parameters like note pitch, volume, onset, and duration from the tracks that correspond to single instruments. If we consider the peculiarities of a musical instrument such as the bass guitar, different playing techniques need to be distinguished to obtain a more realistic parametric representation. The ability to extract these techniques will be beneficial for transcription, expressive performance analysis as well as for genre and artist classification. This paper is organized as follows. After outlining the goals and challenges of this publication and the problems we face in Sect. 2, we provide an overview of related work in Sect. 3. Then, we give a brief explanation of the plucking and expression styles covered in this paper and illustrate the applied post-processing and feature extraction steps in Sect. 4. In Sect. 5, we explain the performed experiments and discuss the obtained results. Finally, Sect. 6 concludes this work.

2. GOALS & CHALLENGES

We aim to design and evaluate different audio features for two classification tasks, namely the classification of plucking and expression styles related to the bass guitar. Concerning their benefit towards classification accuracy, we compare different feature selection, feature space transformation, and classification techniques in the above-mentioned scenarios. Furthermore, we introduce a novel instrument sample database as a public benchmark for these two classification tasks. The main challenges are the high demands on the spectral estimation due to the short attack part of a note played on a plucked string instrument as well as the design of the domain-specific audio features to capture the known peculiarities of the sound-production associated to each style.

3. PREVIOUS APPROACHES

Existing bass transcription algorithms were proposed amongst others in [1] and [2]. Some of the playing techniques investigated in this paper have been studied separately for the purpose of sound synthesis for the guitar in different publications. Flageolet tones (see harmonics, Sec. 4.1) were covered for instance in [3], vibrato in [4].

Many publications dealt with the model-based synthesis of guitar notes [5]. They emphasized the importance of a suitable excitation function to be used as the model input signal. These functions are assumed to correspond to different plucking styles and thus were extracted by an inverse filtering of recorded guitar notes using the synthesis model. A digital waveguide model for the synthesis of a slapped bass guitar considering the physical conditions during the sound production was introduced in [6] (see the slap techniques, Sec. 4.1).

4. NEW APPROACH

4.1. Plucking and expression styles

Depending on whether a particular style is executed by the plucking hand or the gripping hand, we discern plucking styles and expression styles. In this paper, we distinguish between the 5 plucking styles finger-style (FS), picked (PK), muted (MU), slap-thumb (ST), and slap-pluck (SP) and the 5
expression styles normal (NO), vibrato (VI), bending (BE), harmonics (HA), and dead-note (DN).

Finger-style (FS) usually describes the alternating use of the index and middle finger of the playing hand. Picked (PK) characterizes the plucking of the string using a plastic pick instead of the fingers. Slap-thumb (ST) and slap-pluck (SP) describe striking of a string using the thumb and the picking of a string using either the index or the middle finger. Both techniques cause the string to hit the higher frets of the instrument due to its high decaying factor (SCM) [10] to measure the flatness within the PSD of a given time frame $\tau$. The two slap bass techniques cause the string to hit the higher frets of the instrument due to its high damping to better detect the plucking style MU.

The modified covariance method has been shown to provide the best spectral resolution for analyzing sinusoidal signals among other AR methods such as the Yule-Walker, Burg, and covariance method [8]. The filter parameters $a_k(\tau)$ are derived from previously estimated forward and backward linear prediction coefficients by minimizing their prediction squared errors. See [8] for more details on the modified covariance method. We used a block-size of 128, a hop-size of 16, and an AR model order $p = 80$. Previous to the spectral estimation, the time signal is down-sampled to $f_s = 10.025$ kHz.

4.3. Post-processing

We derive the envelope function $X_k(\tau)$ of the first 20 partials by a frame-wise tracking of the harmonic peaks within the estimated PSD. The frequency path of the $k$-th partial is denoted as $f_k(\tau)$. The 0-th partial corresponds to the fundamental frequency $f_0$. We use a two-state envelope model as depicted in Fig. 1 that consists of an attack part $(\tau_1,k \leq \tau \leq \tau_2,k)$ and a decay part $(\tau_2,k < \tau \leq \tau_3,k)$ to approximate the envelope function $X_k(\tau)$ of each partial $k$. We detect the onset times $\tau_1,k$ and offset times $\tau_3,k$ by applying an amplitude threshold of 5% of the maximum amplitudes at $\tau_2,k$.

4.4. Feature extraction

In this subsection, we present the audio features that we applied for given classification tasks. If not otherwise stated, we derive the statistical descriptors minimum, maximum, median, mode, variance, kurtosis and skewness from all time-dependent features described in this section. This allows to characterize the course of different features over time in the attack and the decay part separately.

Features motivated by plucking styles

To simplify the envelope model, we assume that the envelope function $X_k(\tau)$ of each partial $k$ can be approximated by a linear increasing function such as $X_{lin,k}(\tau) = a_k \tau + b_k$ during its attack part and by a decaying exponential function such as $X_{exp,k}(\tau) = c_k \exp(-d_k \tau)$ during its decay part. We use $a_k = [X_k(\tau_2,k) - X_k(\tau_1,k)] / (\tau_2,k - \tau_1,k)$ and $d_k = \ln X_k(\tau_2,k) - \ln X_k(\tau_3,k) / (\tau_3,k - \tau_2,k)$ as features to characterize the envelope shape of each partial $k$. This way, we can distinguish between different intensities of string damping to better detect the plucking style MU.

To characterize the percussiveness of a note and thus to detect the non-tonal dead-notes (DN), we use the spectral crest factor (SCM) [10] to measure the flatness within the PSD of a given time-frame $\tau$. The two slap bass techniques
SP and ST are characterized by a percussive attack and a
tonal decay part. To capture this, we derive the aforementioned statistical
descriptors from the time derivative of the
SCF over all frames of the attack part to characterize the
temporal evolution of the sound characteristic. As described
in Sec. 4.1, both slap styles (SP and ST) result in a bright and
metal-like sound. The spectral centroid [10] characterizes a
sound to either have a bright or dark characteristic.

Features motivated by expression styles
If harmonics (HA) are played on the electric bass, the spectral
energy is mainly distributed towards higher order partials.
Thus, we calculate the partial presence values as the ratios
between the maximum magnitude values $X_k(\tau_{2,k})$ of each
partial and the maximum magnitude value $X_0(\tau_{2,0})$ of the
partial that corresponds to the fundamental frequency $f_0$ as
features. Moreover, we take the three tristimulus values,
the spectral irregularity, and the spectral brightness [11] to
obtain different characterizations of the energy distribution
over all partials for each frame. As additional spectral
descriptors of the attack and decay part, we derive statistical
features from the course of the spectral decrease, the
spectral skewness, the spectral slope and the spectral spread [10]
separately for both parts.

The two styles BE and VI are characterized by typical fre-
quency modulations during the decay part of a note. We apply
a 512 point Fast Fourier Transform (FFT) on the frequency
path $f_k(\tau)$ (of each partial over all frames of the decay part)
resulting in its Fourier transform $F_k(u)$ with $u$ being the
modulation frequency of the partial frequency in Hz. We
look for the maximum value of $|F_k(u)|$ within the modulation
frequency range $4 \leq u \leq 20$, which we found to be the typical
range for vibrato (VI) and bendings (BE) in the applied
data set. Therefore, we take both the detected peak frequency $u_{\text{max}} = \arg \max_u |F_k(u)|$ as well as the difference $|F_k(u_{\text{max}})| - |F_k(u)|$ between the maximum and the mean
value within this frequency range to measure the dominant
modulation frequency and the overall intensity of the modula-
tion. To distinguish between bending and vibrato, we extract
the number of periods within the frequency path $f_k(\tau)$ during
the decay part. We subtract the mean of $f_k(\tau)$ and estimate
the number of half-waves by using a simple sign threshold
criterion. Only segments with a constant sign that are at least
60% as long as the longest segment are taken into account.
The number of half-waves are taken as feature.

5. EXPERIMENTS AND RESULTS

We performed two experiments namely the separate classi-
fication of the 5 plucking styles and the 5 expression styles
introduced in Sec. 4.1. For reasons of simplification, we
assume the expression style normal (NO) for the classification
of the different plucking styles and the plucking style fingerstyle (FS) for the classification of the different expression
styles in the two experiments (see Sec. 5.3). We performed a
baseline experiment for both styles separately using Mel Fre-
quency Cepstral Coefficients (MFCC) as features and Gaus-
sonian Mixture Model (GMM) classifiers with a varying number
of gaussians $n$ between 1, 2, 3, 5, and 10. We achieved best
classification accuracies of 65.7% ($n = 2$) and 67.3% ($n = 3$)
for the classification of plucking and expression styles.

5.1. Experimental Setup

We assembled a novel data set consisting of recorded notes in-
cluding all 10 aforementioned plucking and expression styles.
The applied combinations between playing and expression
styles correspond to the experiments explained in Sec. 5,
the data set will be extended using other combinations in the fu-
ture. In this paper, we only used isolated notes to avoid over-
lapping note segments. We used 3 different bass guitars each
with 3 different pick-up settings to cover a wide timbral range
of instrument sounds. Using the most common pitch range
of a 4-string bass guitar between from E1 (41.2 Hz) to G3
(196.0 Hz), about 4300 notes have been recorded covering all
10 styles. We intend this data set to be a public benchmark
set for the given tasks\footnote{See http://www.idmt.fraunhofer.de/eng/business%20areas/dataset_bass_guitar.htm
for further information.}.

5.2. Feature selection (FS), feature space transformation
(FST) and classification

Overall, we obtain a 224-dimensional feature vector. We in-
vestigated the feature selection technique Inertia Ratio Maxi-
mization using Feature Space Projection (IRMFS) as well as the
feature space transformation techniques Linear Discrimi-
nant Analysis (LDA) and Generalized Discriminant Analysis
(GDA) as pre-processing steps to reduce the dimensionality
of the feature space for an improvement of the subsequent
classification. As classifiers, we compared Support Vector
Machines (SVM) with a Radial Basis Function (RBF) kernel,
GMM, Naive Bayes (NB), and $k$-Nearest Neighbor (kNN).
More details on these methods can be found in [12] and [13].

5.3. Results

We perform multiple classification runs for plucking styles
and expression styles to derive the mean and standard devi-
ation of the classification accuracy. Therefore, the data-set
was partitioned into training set and test set according to a ra-
tio of 90% and 10%. A 10-fold cross-validation was applied.
The best classification results for all investigated classifiers
are depicted in Tab. 1 for both experiments. The table also
contains the parameters $n$ as the number of Gaussians (GMM
classifier), $k$ as the number of nearest neighbors (kNN clas-
sifier), $d$ as the number of selected features (IRMFS feature
selection) and $\gamma$ as another model parameter (for GDA). See
[13] for details. We achieved best mean classification scores
of 93.25% and 95.61% for the classification of plucking and expression styles. The combination of IRMFSP for feature selection and GDA for feature space transformation lead to the highest classification scores for most of the classifiers. Note, that the nonlinear FS and FST methods make classification problem easier linearly solvable and minimize the influence of the classifier itself, so that even simple classifiers like NB perform comparably to non-linear ones like SVM.

6. CONCLUSIONS

In this paper, we introduced a set of low-level features that allow to model the peculiarities of 10 different bass-related plucking and expression styles by capturing typical timbre-related characteristics. We compared 4 different classifiers in combination with one feature selection and two feature space transformation algorithms within two classification tasks. A novel database of isolated bass notes was assembled for evaluation purpose. It is intended as an open benchmark for the given tasks. According to the results of the baseline experiments, the application of more domain-specific features for this task has been shown to increase classification accuracy by 27.55% and 28.31% points up to 93.25% and 95.61%.

7. REFERENCES


Table 1. Mean classification accuracy values [%] (standard deviation [%] given in brackets) for different classifiers without and with feature selection (FS) / feature space transformation (FST). (further parameters explained in Sec. 5.3)