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.

This work has been partly supported by the German research project *GlobalMusic2One* funded by the Federal Ministry of Education and Research (BMBF-FKZ: 01/S08039B) (http://www.globalmusic2one.de).

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

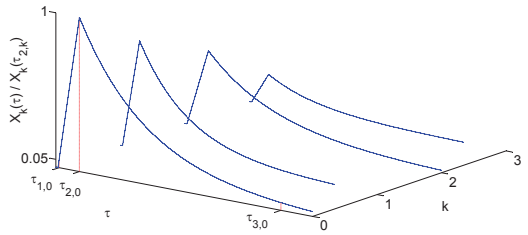

Fig. 1. Applied envelope model for the first 4 partials (normalized magnitude)

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 deflection which results in a metal-like sound. *Muted* (MU) describes plucking the string using the thumb of the plucking hand while simultaneously damping the vibrating string by using the inner side of the plucking hand. Once the string is plucked, the gripping hand can remain on a particular fret (NO) or change the string pitch by bending the string consecutively upwards and downwards once (BE) or periodically (VI). If the string is prevented from vibrating but damped by using the gripping hand, a percussive note sound (DN) emerges. *Harmonics* (HA) can be played by softly damping the string on positions that correspond to integer fractions of the string length. This results in the sounding of the respective harmonics.

4.2. Spectral estimation

Plucking styles related to string instruments like the bass guitar mainly affect the attack period of a note, which usually only lasts up to approximately 40 ms [7]. To provide a sufficient time and frequency resolution, techniques such as auto-regressive modeling, Wigner-Wille Transform, or Wavelet Transform are applied instead of the commonly used short-time Fourier Transform.

In this paper, we use the *modified covariance method* for the parametric estimation of the power spectral density (PSD). It is an auto-regressive (AR) spectral estimation method that has shown to be well suited for the analysis of the short attack transients [8] [9]. Classical spectral estimation methods are based on a windowing of the auto-correlation sequence (ASC) in the time domain during the estimation of the PSD. This causes the side-lobe phenomenon and thus limits the spectral resolution. In contrast, during the AR spectral estimation methods, the ACS values are reconstructed outside the observation interval [8]. The magnitude of the PSD $P_{xx}(f, \tau)$ is modeled by an all-pole transfer function of an AR filter such as $H(z) = 1/(1 - \sum_{k=1}^{p} a_k(\tau) z^{-k})$ for each time frame τ. AR filter transfer functions have sharp peaks and are well-suited to model the sparse spectrum of a harmonic sound. 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 a AR model order of $p = 80$. 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 τ. 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 frequency 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 $u_{max} = \arg\max_u |F_k(u)|$ as well as the difference $|F_k(u_{max})| - \overline{|F_k(u)|}$ between the maximum and the mean value within this frequency range to measure the dominant modulation frequency and the overall intensity of the modulation. 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 classification 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 Frequency Cepstral Coefficients (MFCC) as features and Gaussian 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 including 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 future. In this paper, we only used isolated notes to avoid overlapping 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[1].

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

Overall, we obtain a 224-dimensional feature vector. We investigated the *feature selection* technique Inertia Ratio Maximization using Feature Space Projection (IRMFSP) as well as the *feature space transformation* techniques Linear Discriminant 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 deviation of the classification accuracy. Therefore, the data-set was partitioned into training set and test set according to a ratio 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 classifier), d as the number of selected features (IRMFSP feature selection) and γ as another model parameter (for GDA). See [13] for details. We achieved best mean classification scores

[1] See http://www.idmt.fraunhofer.de/eng/business%20areas/dataset_bass_guitar.htm for further information.

	Plucking Styles			Expression styles		
	Without FS / FST	Best results with FS / FST		Without FS / FST	Best results with FS / FST	
Classifier	**Acc.: Mean (Std.)**	**Acc.: Mean (Std.)**	**Best FS / FST configuration**	**Acc.: Mean (Std.)**	**Acc.: Mean (Std.)**	**Best FS / FST configuration**
SVM	90.75% (0.39%)	92.77% (0.55%)	IRMFSP(80) + GDA (10^{-7})	93.77% (1.20%)	94.96% (0.29%)	IRMFSP(100) + GDA(10^{-7})
GMM(2)	70.04% (0.06%)	92.30% (1.23%)	IRMFSP(80) + GDA (10^{-7})	77.63% (3.52%)	95.13% (1.44%)	No FS + GDA(10^{-15})
GMM(3)	72.61% (3.96%)	92.32% (0.59%)	IRMFSP(80) + GDA (10^{-7})	75.09% (1.20%)	94.78% (0.71%)	No FS + GDA(10^{-9})
GMM(5)	75.92% (0.39%)	**93.25% (0.12%)**	**IRMFSP(100) + GDA**(10^{-7})	77.62% (2.05%)	95.28% (0.23%)	IRMFSP(100) + GDA(10^{-7})
GMM(10)	79.22% (1.59%)	92.51% (1.53%)	IRMFSP(80) + GDA (10^{-7})	82.45% (4.10%)	**95.61% (0.80%)**	**IRMFSP (100) + GDA**(10^{-7})
NB	66.43% (5.72%)	91.63% (0.96%)	IRMFSP(80) + GDA(10^{-7})	72.61% (3.65%)	95.28% (0.23%)	IRMFSP(100) + GDA(10^{-7})
kNN(1)	79.62% (3.22%)	92.34% (0.05%)	IRMFSP(100) + GDA (10^{-7})	87.35% (0.75%)	94.79% (0.57%)	IRMFSP(100) + GDA(10^{-7})
kNN(5)	82.58% (3.76%)	92.96% (1.49%)	IRMFSP(80) + GDA (10^{-7})	90.08% (0.91%)	95.12% (0.45%)	IRMFSP(100) + GDA(10^{-7})
kNN(10)	82.61% (1.01%)	92.80% (0.09%)	IRMFSP(100) + GDA (10^{-7})	90.45% (0.67%)	94.97% (0.65%)	IRMFSP(100) + GDA(10^{-7})

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)

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

[1] C. Dittmar, K. Dressler, and K. Rosenbauer, "A toolbox for automatic transcription of polyphonic music," in *Proc. of the Audio Mostly conference*, 2007.

[2] M. P. Ryynänen and A. P. Klapuri, "Automatic transcription of melody, bass line, and chords in polyphonic music," *Computer Music Journal*, vol. 32, pp. 72–86, 2008.

[3] J. Pakarinen, "Physical modeling of flagolet tones in string instruments," in *Proc. of the European Signal Proc. Conf. (EUSIPCO)*, 2005.

[4] I. Barbancho, C. de la Bandera, A. M. Barbancho, and L. J. Tardon, "Transcription and expressiveness detection system for violin music," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, 2009, pp. 189–192.

[5] V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, "Discrete-time modelling of musical instruments," *Reports on Progress in Physics*, vol. 69, pp. 1–78, 2006.

[6] E. Rank and G. Kubin, "A waveguide model for slapbass synthesis," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, 1997, pp. 443–446.

[7] R. M. French, *Engineering the Guitar - Theory and Practice*, Springer Science+Business Media, 2009.

[8] S. Lawrence Marple Jr., *Digital Spectral Analysis*, Prentice-Hall, 1987.

[9] F. Keiler, C. Karadogan, U. Zölzer, and A. Schneider, "Analysis of transient musical sounds by auto-regressive modeling," in *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-03), London, UK*, 2003.

[10] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," Tech. Rep., 2004.

[11] K. Jensen, *Timbre Models of Musical Sounds*, Ph.D. thesis, University of Copenhagen, 1999.

[12] J. Abeßer, H. Lukashevich, C. Dittmar, and G. Schuller, "Genre classification using bass-related high-level features and playing styles," in *Proc. of the ISMIR conference*, 2009.

[13] H. Lukashevich, "Feature selection vs. feature space transformation in automatic music genre classification tasks," in *Proc. of the AES Convention*, 2009.