A FEATURE RELEVANCE STUDY FOR GUITAR TONE CLASSIFICATION

Wolfgang Fohl **Ivan Turkalj** **Andreas Meisel**
HAW Hamburg University of Applied Sciences
{wolfgang.fohl | ivan.turkalj | andreas.meisel}@haw-hamburg.de

ABSTRACT

A series of experiments on the automatic classification of classical guitar sounds with support vector machines has been carried out to investigate the relevance of the features and to minimise the feature set for successful classification. Features used for classification were the time series of the partial tone amplitudes, and of the MFCCs, and the energy distribution of the nontonal percussive sound that is produced in the attack phase of the tone. Furthermore the influence of sound parameters as timbre, player, fret position and string number on the recognition rate is investigated. Finally, several nonlinear kernels are compared in their classification performance. It turns out, that a selection of 505 features out of the full feature set of 1155 elements does only reduce the recognition rate of a linear SVM from 82% to 78%. With the use of a polynomial instead of a linear kernel the recognition rate with the reduced feature set can even be increased to 84%.

1. INTRODUCTION

In the recent years musical instrument recognition has been extensively investigated. Primary research goals were automatic indexing of multimedia data bases, automatic musical genre classification and automatic music transcription systems. A less common topic is the quality assessment of musical instruments, which will be covered in the present paper. Currently the research efforts show several trends. One is the attempt to transfer the successful classification based on single notes to the analysis and classification of solo musical phrases. Joder, Essid, and Richard [9] describe a modification of Support Vector Machines with alignment kernels which they report to perform better than classifiers based on Gaussian Mixture Models or Hidden Markov Models. Barbedo and Tsanetakis [2] published their results on the even more challenging task of instrument classification in polyphonic recordings. Their method is the detection of partial tone structures that are unique to certain instrument groups, which are fed to a specialised decision-tree algorithm.

A second trend in instrument classification research is the systematic comparison of the performance of certain statistical methods and commonly used feature sets in order to reduce the feature space dimensionality and thus escaping the "curse of dimensionality" – the exponential increase of required training samples with increasing dimension of the feature space. A detailed general description of algorithms and procedures of feature space reduction is given by Guyon and Elisseeff [6]. Deng, Simmermacher, and Cranefield published a very good and complete survey on feature relevance for musical instrument classification [3]. Loughran, Walker, and O'Neill present a genetic algorithm approach to feature selection [11]. Genetic algorithms can even be employed for feature *generation*, as described by Mierswa and Morik [12], and by Pachet and Roy [13]. One outcome of these approaches could be the generation of meaningful features, that give an insight in the nature of the investigated sounds.

There is a paper on the quality assessment of musical instruments by Hsiao and Su [7]. They used an waveform-based feature set in conjunction with a multiclass Mahalanobis-Taguchi system to develop an automatic saxophone quality assessment system.

In this article we present a systematic study on the parameters influencing the classification performance of a support vector machine (SVM) classifier to distinguish single tones of three different classical guitars from each other. This continues an earlier work published on the 2008 DAFx conference [4].

Our research motivation is to pinpoint those acoustical features of high-quality instruments that are responsible for the perceived musical quality of the guitar. The classification experiments were conducted to further support our working hypothesis, that the acoustical quality of an instrument reveals itself at least partially already in a single tone. This hypothesis is motivated by the way professional guitarists assess an instrument: They test the acoustical properties of the guitar by carefully listening to single tones played on all strings and over the whole range of the fingerboard.

In the following section our experimental setup and the feature and sample selection strategy is described, followed by the presentation and discussion of classification performance results. The article is concluded by a summary and an interpretation of the results in terms of musical acoustics.

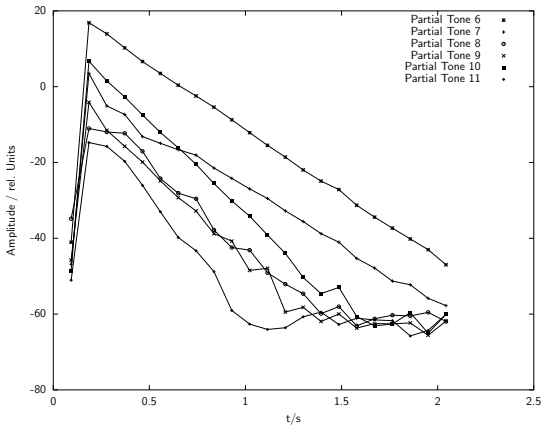

Figure 1. Time series of partial tones 6 – 11 of a guitar tone.

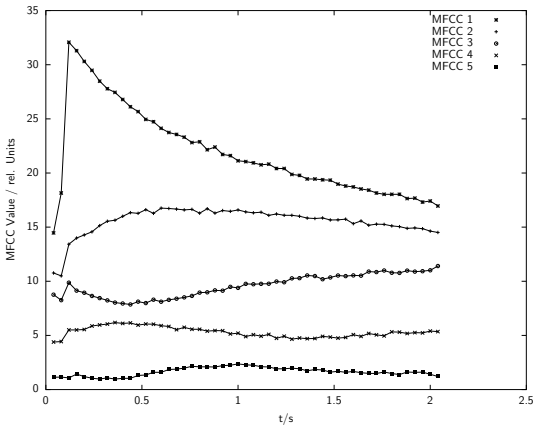

Figure 2. Time series of the first 10 MFCCs of a guitar tone. Curves shifted vertically for better overview.

2. EXPERIMENTS

Single guitar tones of three high-quality classical guitars (by the luthiers Hense, Marin, and Wichmann) played by four players on three different fret positions with the three different sound intentions *sonorous*, *sharp*, and *warm*, were recorded, resulting in ≈ 4000 tone samples, each with a duration of 2–3 seconds. These samples were normalised for equal maximum amplitude and three groups of features were extracted: The time series of the partial tone amplitudes (PT) (see figure 1 for an example), the time series of the mel frequency cepstral coefficients (MFCC) as shown in figure 2, and the power distribution of the *nontonal spectrum* (NT), see [5]. The partial tone time series was obtained by first taking the magnitude spectrum of the whole length of the sound sample, and f_0 was identified by cepstral analysis. Then the sound was split into frames of 4096 samples, each frame multiplied with a Blackman window, the FFT was calculated and the amplitudes of the first 16 partial peaks, i.e., the peaks in the vicinity of $n \cdot f_0$, were evaluated. The data of the first 40 frames was taken as the partial tone feature set.

The MFCC data was computed using the Matlab Auditory Toolbox by Slaney [14], with frame size of 1024 samples, and a frame frequency of 25 Hz. The time series of the first 10 MFCCs was evaluated, and the first 50 frames were taken as the MFCC feature set.

The calculation of the nontonal features starts with the magnitude spectrum of the whole sound sample, from which the tonal peaks have been removed as shown in Fig. 3. To obtain the nontonal features, the nontonal power spectrum P_k at frequency index k is computed by squaring the nontonal spectrum Y_k, and the accumulated power between $f_{\mathrm{start}} = 0$ and $f_{\mathrm{end}} = f_k$ is calculated to yield the fuction C_k:

$$C_k = \sum_{i=0}^{k} P_k \qquad (1)$$

The logarithm of this monotonously increasing function is taken, and the range of $\log C_k$ from 1 to its final value

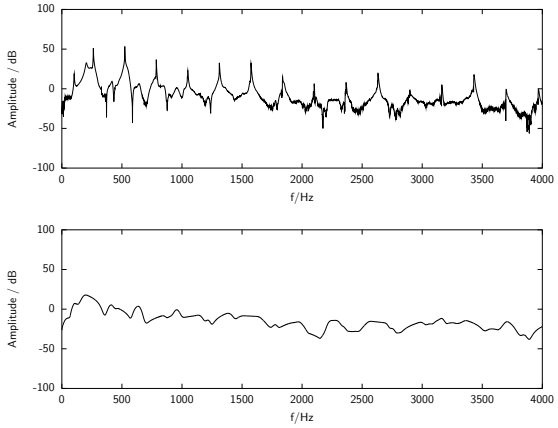

Figure 3. Magnitude spectrum and nontonal magnitude spectrum of a guitar tone.

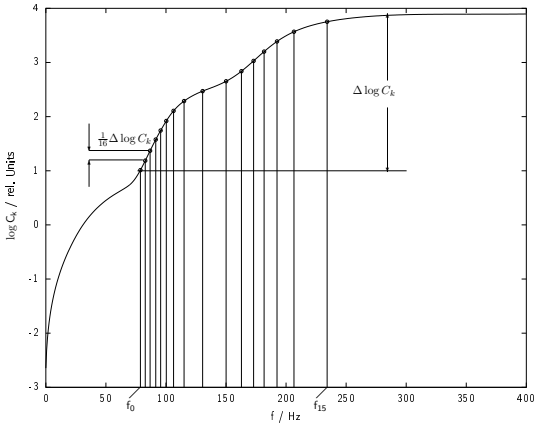

Figure 4. Calculation of nontonal frequency features. Plotted is the function $\log C_k$ from Eq. 1 and the frequencies, that divide the range $\Delta \log C_k$ from $\log C_k = 1$ to the maximum value of $\log C_k$ into 16 equally spaced regions. The corresponding frequencies $f_1 \ldots f_{15}$ are the nontonal feature values.

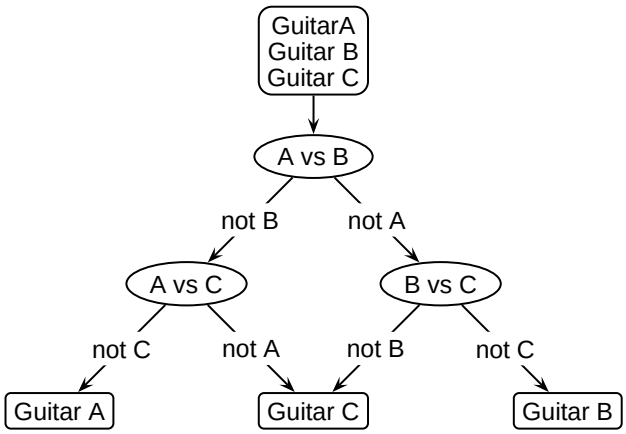

Figure 5. A directed acyclic graph for multi-class identification

is split into 16 equally spaced parts. The corresponding 15 boundary frequencies are taken as the nontonal features, as shown in figure 4.

The complete feature set consists of 1155 elements ($40 \cdot 16\,\mathrm{PT} + 50 \cdot 10\,\mathrm{MFCC} + 15\,\mathrm{NT}$).

These features or subsets of them were used for training the SVMs and for running the classification tests. All but the last group of experiments were performed with linear SVM kernels, the best classification performance with linear kernels was $82.0\,\%$, it was achieved using the full feature set of 1155 elements.

For each guitar, a SVM for a one-vs-rest classification was trained. For the multi-class identification, a directed acyclic graph (DAG) was constructed according to Fig. 5.

The *classification performance* is determined as the ratio of correctly classified test examples and the total number of examples

$$Performance = \frac{n_{\mathrm{correct}}}{n_{\mathrm{total}}} \tag{2}$$

A flexible data processing software has been written in GNU Octave [1], a free Matlab™ clone, for feature extraction and data selection. For the SVMs, the SVMLight implementation of Joachims has been used [8].

Several series of experiments were carried out to investigate which features are most important for a correct classification, how small the feature set can be made without degrading the classification performance, and if there are nonlinear kernels that perform better than the linear kernel.

In addition, the training and testing was conducted with different subsets of the guitar tone samples to investigate the role of the player, and to test, if a preselection of sounds according to several criteria can improve the classification result.

Two preliminary experiments were performed: A cross-validation of the classification was made by exchanging the sample sets for test and training. In the second preliminary experiment an attempt was made to reduce the feature space by performing a Principal Component Analysis on the training data set.

3. RESULTS AND DISCUSSION

3.1 Classification Performance With the Full Feature Set and a Linear Kernel

As a first step, the classification experiment has been carried out with the full data set of 1155 features, 978 training samples, 972 test samples, with a linear kernel.

The classification performance of $82.0\,\%$ is taken as the reference value for all subsequent experiments.

3.1.1 Cross-validation

The experiment was repeated with the test and training data exchanged. Table 1 compares the results of the two experiments.

Guitar	*Reference*	*Test / Train Data Exchanged*
Overall	$82.0\,\%$	$84.2\,\%$
Hense	$74.4\,\%$	$82.9\,\%$
Marin	$81.8\,\%$	$77.7\,\%$
Wichmann	$88.6\,\%$	$92.0\,\%$

Table 1. Cross-validation of guitar classification with exchanged train and test data

The deviations between the two experiments give an impression of the variances of the classification measurements.

3.2 Principal component analysis (PCA)

As a supporting study, an attempt was made to reduce the dimensionality of the feature space by applying a principal component analysis to the feature data. With the first 500 PCA-Eigenfeatures the classification rate is only $72.1\,\%$, which is significantly worse than the classification result of $77.7\,\%$, based on a manually selected 505-feature set shown in section 3.3.3. An explanation for this poor result might be the fact, that most of the features are *time series*, for which the differences of the neighbouring values carry important information. The preprocessing in the `princomp` function of the Octave/Matlab Statistical Toolbox removes these dependencies by calculating the mean and variance for each feature in the training set separately, and then for each feature value subtracts the mean and scales the variance to unity, thus eliminating the information of the relative magnitudes of the feature values.

To get a substantial reduction of the feature space, an adaption of the PCA method for time series, as described in chapter 12 of the book of Jolliffe [10], will have to be applied.

3.3 Relevance of Features

The experiment series to determine the relevant features were carried out with the full sample data set: 978 samples for training, and 972 samples for testing.

3.3.1 MFCCs

In this series of experiments only the time series of MFCC features were used for classification. In the first experiment set the MFCC coefficients were grouped into lower and upper half (coefficients 1 – 5 and 6 – 10), in the second experiment set the features were divided into three groups (coefficients 1 – 3, 4 – 7, and 8 – 10). All possible combinations of groups in the series were tested, these are the most important results:

MFCC coefficients	*Features*	*Performance*
1 – 10	500	75.5 %
1 – 5	250	71.2 %
6 – 10	250	60.2 %
1 – 7	350	72.0 %
1 – 3	150	65.9 %
4 – 7	150	62.6 %
8 – 10	150	56.8 %

Table 2. Classification performance of subsets of the MFCC features

The MFCC coefficient group 1 – 3 contains the most relevant third of the MFCC coefficients.

3.3.2 Partial Tones (PT)

In this series of experiments several selections of the first 16 partial tones have been made. Again a grouping in halves and thirds has been performed.

Partials	*Features*	*Performance*
1 – 16	640	52.8 %
1 – 11	440	57.2 %
1 – 5	200	44.0 %
6 – 11	240	54.3 % (!)
12 – 16	200	42.7 %

Table 3. Classification performance of subsets of the PT features

The medium third of the partial tones gives not only the best result of the one-third-selection, it is noteworthy, that this reduced feature set even performs better than the full set of partial tones.

3.3.3 Feature Combinations

Several combinations of the nontonal, MFCC, and partial tone features were tested. The best performance was achieved by combining nontonal features with MFCCs 1 – 5 and partial tones 6 – 11 i.e., the best-performing selections of the previous experiments:

Comparison of the first two lines in Table 4 shows, that the addition of the 15 nontonal features increases the classification performance by 5 percent points. It has to be stressed, that the MFCC and the PT features each represent a time series of 50 (MFCC) and 40 (PT) data points respectively, whereas the nontonal energy distribution is a global feature set that consists of only 15 single data values. So it can be concluded, that the nontonal features add new information to the feature set, that is not implicitly contained in the other feature data.

Feature selection	*Features*	*Performance*
MFCC 1 – 5, PT 6 – 11 NT 1 – 15	505	77.7 % (!)
MFCC 1 – 5, PT 6 – 11	490	71.9 %
MFCC 1 – 5, NT 1 – 15	265	76.0 %
MFCC 1 – 7, PT 6 – 11 NT 1 – 15	605	77.7 %

Table 4. Classification performance of various feature combinations

Another notable fact is, that the inclusion of the time series of MFCCs 6 and 7 does not at all affect the classification performance, as can be seen by comparing the first and the last line of Table 4.

3.4 Selections of Tone Samples

In the subsequent experiments certain selections of tone samples were made to further pinpoint the relevance of features. Since the number of training samples is reduced, also a reduced parameter set has to be used, so the most successful combination of the preceding experiments was taken: nontonal features 1 – 15, MFCCs 1 – 5, and partial tones 6 – 11. This is the 505-element feature set of section 3.3.3. In each overview of results the number of audio samples used for test and training is given.

3.4.1 Player

In the first set of the experiments with preselected tone samples, the influence of the player is investigated. In the first part, three out of four players are used for training, the remaining player is taken for testing. In the second part, the training is performed with all players, and again one of the players is used for testing.

Player (Test)	*Samples (Test)*	*Players (Train)*	*Samples (Train)*	*Performance*
1	324	2, 3, 4	654	59.6 %
2	162	1, 3, 4	816	69.1 %
3	324	1, 2, 4	648	65.4 %
4	162	1, 2, 3	816	64.2 %

Table 5. Classification performance, player in test set not included in training set

Player (Test)	*Samples (Test)*	*Players (Train)*	*Samples (Train)*	*Performance*
1	324	1 – 4	978	74.4 %
2	162	1 – 4	978	71.6 %
3	324	1 – 4	978	83.5 %
4	162	1 – 4	978	78.4 %

Table 6. Classification performance, player in test set **is** included in training set

As was to be expected, the players have quite a large influence on the produced sound, and so the classification

performance decreases, when the testing player is not in the group of the training players. The classification performance might in this case be improved by a larger pool of players.

3.4.2 Timbre

In these experiments sound samples of the same timbre are used for training and testing. The term timbre here refers to the sound intention of the player. Usually, a warm timbre is produced by plucking the string above the sound hole of the guitar with the finger moving in an angle of approx. $45°$ to the string; a sharp timbre is produced by plucking near the bridge with the plucking finger moving perpendicular to the string.

Timbre	*Samples (Train / Test)*	*Performance*
sharp	326 / 324	80.6 %
sonorous	327 / 324	79.0 %
warm	325 / 324	82.7 %

Table 7. Classification performance for different timbres

It would have been expected, that the preselection of timbre would improve the classification performance, but the experiments show, that this is not the case. Obviously the influence of the different strings and the different positions on the fingerboard introduce too much inhomogeneity.

3.4.3 String

This series of experiments tests the influence of the string on the sound. Only sounds of the same string are taken for training and testing.

String (Note)	*Samples (Train / Test)*	*Performance*
1 (e')	165 / 162	96.3 %
2 (b)	163 / 162	92.6 %
3 (g)	162 / 162	80.9 %
4 (d)	163 / 162	79.6 %
5 (A)	163 / 162	84.0 %
6 (E)	162 / 162	83.3 %

Table 8. Classification performance for different strings

Obviously the preselection of the string does provide substantially more homogeneous sample sets. The trained SVMs are specialised to the sound of one particular string and perform substantially better than with the whole range of tone samples.

3.4.4 Fret

The last experiment series in this sections is devoted to the fret, i.e., the position on the fingerboard .

The observed performances of Table 9 are approximately the same as the overall performance given in Table 1.

Fret	*Samples (Train / Test)*	*Performance*
1	326 / 324	84.9 %
5	176 / 174	79.9 %
6	150 / 150	78.7 %
10	314 / 312	81.7 %

Table 9. Classification performance for different fret positions

4. NONLINEAR KERNELS

In a last series of experiments different nonlinear kernels were used for classification. Again the 505-element feature set of section 3.3.3 is used.

The only nonlinear kernel provided by SVMLight, that gave satisfactory results was the polynomial kernel. Table 10 shows the classification results for several polynomial degrees:

Polynomial Degree	*Samples (Train / Test)*	*Performance*
1	978 / 972	77.7 %
2	978 / 972	82.3 %
3	978 / 972	84.0 %

Table 10. Classification performance for polynominal kernels of degree 1–3

Other available nonlinear kernels (Sigmoid, RBF) and higher degree polynomial kernels performed very poor.

5. SUMMARY AND OUTLOOK

In this paper a detailed feature relevance study about the classification performance of SVMs for classical guitar sounds is presented. It is shown, that the original feature set of 1155 features with a classification performance of 82.0 % can be reduced to 505 features with an even better performance of 84.0 % when employing a third degree polynomial kernel.

Several experiments on the preselection of sound samples for testing and training have been carried out. A tentative interpretation in musical terms shall be tried in the following paragraphs.

The group of experiments with a pool of players used for training of the SVMs and one player for testing shows, that there is a large influence of the player on the sound. This conclusion can be drawn from the fact, that the classification performance is significantly increased, when the testing player is also member of the training players. From musical experience this is plausible. It is the interaction of player and instrument that produces the sound.

The preselection experiments, where the same timbre, string, and fret is used for training and testing can be explained in a technical way: the more uniform the samples are, the easier is the detection of differences arising from the acoustical properties of the guitars. The very good classification performance for the highest two strings (96.3 %) and 92.6 % is in accordance with the experience of luthiers

and guitar players: A good guitar reveals its quality on the treble strings (b and e'), whereas even medium quality guitars may sound good on the lower strings.

It would be a promising approach to improve the overall classification result by introducing a two-step classification process: in the first step the string is determined, and in the second step the guitar is identified. Currently there are experiments going on to compare the reported results with other classification methods, in particular neural networks and a specialised form of principal component analysis to the classification problem presented by Wells and Aldam in [15].

The classification framework is currently being modified to apply to pieces of polyphonic solo guitar music. The robustness of the method has to be proven, when there is a mixture of tones to be analysed, and further features may show up in the musical context, that are not present in the single note analysis, especially the range of possible variations in amplitude, attack and decay times and various spectral properties.

6. REFERENCES

[1] Gnu octave. http://www.octave.org, visited 2011.

[2] J. G.A Barbedo and G. Tzanetakis. Instrument identification in polyphonic music signals based on individual partials. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 401 – 404, 2010.

[3] Jeremiah D. Deng, Christian Simmermacher, and Stephen Cranefield. A Study on Feature Analysis for Musical Instrument Classification. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 38(2):429–438, 2008.

[4] K. Dosenbach, W. Fohl, and A. Meisel. Identification of Individual Guitar Sounds by Support Vector Machines. In *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, 2008.

[5] Dimitrios Fragoulis, Constantin Papaodysseus, Mihalis Exarhos, George Roussopoulos, Thanasis Panagopoulos, and Dimitrios Kamarotos. Automated classification of piano-guitar notes. *IEEE Trans. on Audio, Speech, and Language Processing*, 14(3), 2006.

[6] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.

[7] Y. H Hsiao and C. T Su. Multiclass MTS for saxophone timbre quality inspection using waveform-shape-based features. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(3):690 – 704, 2009.

[8] Thorsten Joachims. Svmlight. http://svmlight.joachims.org, 2008.

[9] Cyril Joder, Slim Essid, and Gaël Richard. Alignment kernels for audio classification with application to music instrument recognition. In *Proceedings of the European Signal Processing Conference*, 2008.

[10] I.T. Jolliffe. *Principal Component Analysis*, volume 2. Wiley Online Library, 2002.

[11] R. Loughran, J. Walker, and M. O'Neill. An exploration of genetic algorithms for efficient musical instrument identification. In *Signals and Systems Conference (ISSC 2009), IET Irish*, pages 1–6. IET, 2009.

[12] Ingo Mierswa and Katharina Morik. Automatic feature extraction for classifying audio data. *Machine Learning*, 58(2-3):127–149, 2005.

[13] F. Pachet and P Roy. Analytical features: a knowledge-based approach to audio feature generation. *EURASIP Journal on Audio, Speech, and Music Processing*, 2009(1), February 2009.

[14] Malcolm Slaney. Auditory toolbox. A MATLAB toolbox for auditory modeling work. version 2. http://cobweb.ecn.purdue.edu/ malcolm/interval/1998-010/, 1998.

[15] Jeremy J. Wells and Gregory Aldam. Principal component analysis of rasterised audio for crosssynthesis. In *Proc. Of the 12th Int. Conference on Digital Audio Effects (DAFx-09)*, volume 4, Como, Italy, 2009.