IDENTIFICATION OF INDIVIDUAL GUITAR SOUNDS BY SUPPORT VECTOR MACHINES

Kerstin Dosenbach, Wolfgang Fohl, Andreas Meisel,*

Dept. of Computer Science
University of Applied Science, Faculty TI
Hamburg, Germany
{dosenb_k|fohl|meisel}@informatik.haw-hamburg.de

ABSTRACT

This paper introduces an automatic classification system for the identification of individual classical guitars by single notes played on these guitars. The classification is performed by Support Vector Machines (SVM) that have been trained with the features of the single notes. The features used for classification were the time series of the partial tones, the time series of the MFCCs (Mel Frequency Cepstral Coefficients), and the "nontonal" contributions to the spectrum. The influences of these features on the classification success are reported. With this system, 80% of the sounds recorded with three different guitars were classified correctly. A supplementary classification experiment was carried out with human listeners resulting in a rate of 65% of correct classifications.

1. INTRODUCTION

This paper is part of a research program to determine an *acoustic fingerprint* of an instrument, i.e., a data set that reflects the relevant acoustic parameters making up the "personality" of the instrument. Such an acoustic fingerprint can be utilized for the verification of a physical model, the quality of a synthesizer, or the quality of a real guitar, where the acoustic fingerprint of the generated sound will be compared to the acoustic fingerprint of an reference instrument. The initial question was, whether one single tone already reveals the specific acoustic character of the instrument.

A machine learning system is developed that is able to distinguish between different classical guitars by features extracted from single notes played on these guitars. The main problem in developing such a learning system is, that the player is able to vary the character of the sounds drastically by using different plucking techniques, by varying the location where the string is plucked or by varying the angle between the line of motion of the finger and the string [1]. A proper set of acoustical features has to be determined to identify an individual instrument independent of plucking techniques. Support Vector Machines are taken as machine learning system, and the role of different feature sets on the classification accuracy is investigated.

It will be shown, that this system is able to identify an instrument by one single tone. Obviously this tone bears enough information about the individual instrument to permit the identification.

In the remainder of this section, an overview over related work is given. The second section presents our experimental setup, the third section gives a very short introduction to the application of Support Vector Machines to multi-class identification problems. The results of our measurements are presented in the fourth section. Conclusions and acknowledgements complete the paper.

* Present e-mail address: kerstin.dosenbach@rsmg.de

1.1. Related Work

Classical guitar sounds have been investigated in the past with several methods. There are attempts to develop physical models of the guitar in order to assist the luthier in improving the design of the guitars. [2], [3], or to predict the acoustic properties of novel materials [4]. Another strong motivation for research is the development of realistic synthesizers for classical guitar sounds [5], [6]. A third motivation is the search for effective data compression algorithms without destroying the quality of the sound [5].

The selection of features for classical instrument recognition is described in many papers. Common to all of these works is the focus on identifying the *type* of an instrument (e.g. is it a trombone?) or the *instrument family* (e.g. is it a brass instrument?). This kind of classification is applied in automatic music transcription systems.

Steelant et al. [7] investigate percussive sounds, i.e. sounds produced by the instruments of a drum set. Features were Zero Crossing Rate, Crest Factor, Temporal Centroid, several central momenta of the spectrum. These features are well suited for percussive sounds, but not for tonal sounds.

Marques and Moreno [8] compare Support Vector Machines with Gaussian Mixture Models for the identification of instruments like bagpipes, clarinet, flute, harpsichord, organ, piano, trombone, and violin. They study various feature sets (LPC, Cepstral Coefficients, MFCCs) to identify the instruments and report that SVMs with an MFCC feature set provide the best classification results.

Deng et al. [9] study the influence of feature selection for the classification of 20 instrument types. Nineteen different features are investigated. Again MFCCs are reported to be the most relevant ones.

An interesting approach to distinguish guitar sounds from piano sounds is described by Fragoulis et al. [10]. The authors describe a method to distinguish these sounds by their *nontonal spectrum*. The nontonal spectrum is the remainder of a FFT spectrum after removing the peaks of the partial tones of the sound. It is dependent of the material and geometry of the instrument, but largely independent of the particular tone played.

2. EXPERIMENTAL SETUP

2.1. Recording

To achieve a database of classical guitar sounds, we proceeded as follows: A collection of guitar notes was recorded by 5 players playing 3 guitars of the luthiers Dieter Hense (guitar A), Paco Santiago Marin (guitar B), and Michael Wichmann (guitar C). On each guitar, single notes were played on the 1^{st}, 6^{th} and 10^{th} fret of each string. Each note was played with three different *musical* *timbres*: "warm", "sonorous", and "sharp", and was repeated several times. The strings not being played were carefully damped, and there was no vibrato applied to the tone.

The sound samples were recorded in an anechoic chamber with a condenser microphone AKG C3000 B. The microphone was located 30 cm above the soundboard of the guitar, 10 cm below the bridge. To compensate for near-field-effects, the microphone's built-in 20dB/decade low-pass filter with a cutoff frequency of 500 Hz was activated. The microphone signals were digitized with a MOTU 828mkII audio interface with 24 bit resolution and 44100 Hz sampling rate.

Sound files were produced from the recording sessions, each file containing one single tone. The sound files have a fixed length of 2 seconds. The starting point of the sounds is 0.1 seconds before the maximum amplitude of the sound is reached. The maximum amplitude of each sound file has been normalized to -3 dB, and the information of the original amplitude was stored together with the sound data. The complete sound database is publicly available [1]. Two subsets of the database were taken for training and for testing the SVMs.

A supplementary experiment was carried out to compare the results of the automatic classifications with the human classification capabilities. An audio CD was prepared with three sets of four tones. Each set consisted of one unlabeled tone followed by three tones labelled with the guitar names. The CD was given to 12 test listeners, mostly amateur musicians, who should assign the proper guitar name to the unlabelled tone. They were allowed to repeat the sounds as often as they liked.

2.2. Feature Extraction

All data evaluation algorithms are implemented in Octave / MATLAB

First of all the fundamental frequency f_0 is calculated by cepstrum analysis and introduced to the feature vector. The introduction of the fundamental frequency turned out to be necessary, because the sound quality of any guitar is dependent of f_0 (i.e. the pitch of the tone), and none of the other feature values contains explicit pitch information.

The *nontonal amplitude spectrum* is calculated using a MATLAB implementation of the algorithm given by Fragoulis et al. [10]. The data of the first 15 peaks (frequency and amplitude) of these spectra are entered to the feature vector. Figure 1 shows a semilogarithmic plot of the normal and the nontonal spectrum of a guitar sound, indicating the nontonal peaks.

The time series of the first 16 partial tone amplitudes are calculated with a window size of 2048 samples, corresponding to a frame rate of approx. 20 Hz, and entered to the feature vector.

Finally, the time series of the first 10 MFCCs of the sound are calculated with a MATLAB algorithm published in [11] with a frame rate of 25 Hz, resulting in 50 frames for each sound file. Figure 2 shows some time series of the first four MFCC values. From this figure can be seen, that the similarity in MFCC data for two tones played with the same musical timbre is much greater than the similarity of two tones played on the same guitar. This is true for all features, and illustrates the difficulties in properly classifying the sounds.

To estimate the influence of the various features, several feature subsets are used for training and testing the SVMs.

[1] http://guitarsounds.cpt.haw-hamburg.de

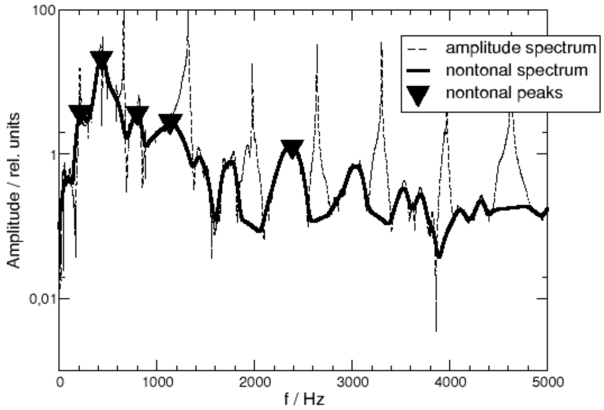

Figure 1: *Normal and nontonal amplitude spectrum*

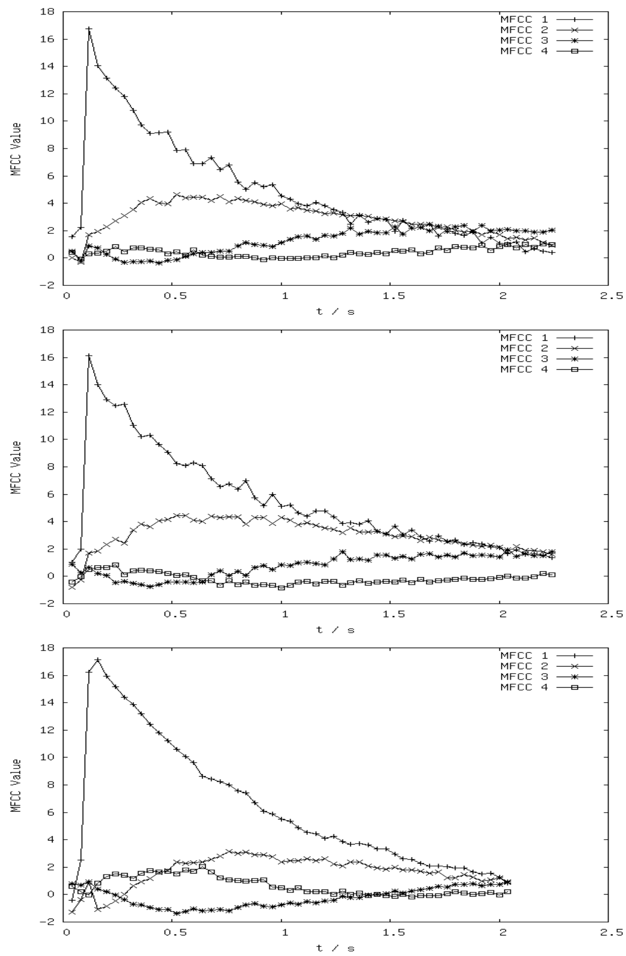

Figure 2: *MFCC data. Top: Guitar A, sonorous. Middle: Guitar B, sonorous. Bottom: Guitar A, sharp*

3. APPLICATION OF SUPPORT VECTOR MACHINES TO MULTI-CLASS IDENTIFICATION

SVMs are machine learning systems for classification. Each data set is represented by a (usually very high-dimensional) *feature vector*. In the training of a SVM an *optimal hyperplane* for the separation of the training data is calculated. The optimal hyperplane is the one with the broadest *margin*, i.e. the largest distance to the feature vectors. A detailed description is given in the book of Schölkopf and Smola [12].

For our investigations the SVM*light* implementation by Joachims [13] is used. This implementation is available as platform-independent source code. SVM*light* is fast and provides detailed debugging output, useful for fine-tuning the learning process.

Support Vector Machines can only separate *two* classes. For multi-class identification problems there are two approaches. The first approach is *one-vs-all-classification*: For each class C_i a SVM $\mathcal{M}_i$ will be trained, that separates the feature vectors of C_i from the feature vectors of all of the other classes. The other approach is a *one-vs-one-classification* where for each *pair* of classes (C_i, C_j) a SVM $\mathcal{M}_{ij}$ is trained that distinguishes between the feature vectors of C_i and C_j. A *Directed Acyclic Graph* (DAG) [14] must be constructed to schedule the necessary comparisons for the classification of an unknown feature vector. This procedure is illustrated in figure 3.

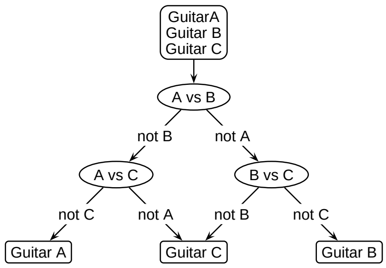

Figure 3: *DAG for multi-class identification*

In the *training phase*, the computational cost for the one-vs-all-classification is lower than for the one-vs-one-classification. In the first case, N SVMs will have to be trained for N classes, in the second case, $N(N-1)/2$ SVMs will have to be trained for each pair of classes. But one-vs-all-classification has the drawback, that the hypersurface dividing the two classes in feature space in most cases will not be a hyper*plane*, requiring the use of modified SVMs (so-called non-linear kernel SVMs). This will increase the computing cost and complicate the configuration of the learning system. For this reason a one-vs-one-classification was used for our experiments.

For the *classification* of an unknown sample the computational costs are nearly equal: N tests for one-vs-all-classification, $N-1$ tests for one-vs-one-classification in conjunction with a DAG (see fig. 3)

4. MEASUREMENTS AND RESULTS

Several feature sets were tested for their influence on the classification performance.

For a detailed view of the classifications, here are the *confusion matrices* for the several feature sets. The row labels indicate the correct guitar, column labels indicate the classification output. Example: the number 34 in row 1, column 2 of the following table means that 34 sound samples of guitar A were erroneously classified as guitar B.

The last confusion matrix shows the result of the human listeners classification experiment.

Abbreviations: PA: Partial tone amplitudes of the first 15 partial tones @ 20 Hz, NT: 15 Nontonal peaks, MFCC: first 10 MFCCs @ 25 Hz

Set 1: f_0, PA, 641 Features, Performance = 48.1%

	Guitar A	Guitar B	Guitar C
Guitar A	**187**	34	103
Guitar B	127	**73**	124
Guitar C	92	24	**208**

For Guitars A and C the majority of classifications is correct, but Guitar B is not properly classified.

Set 2: f_0, NT, 31 Features, Performance = 32.4%

	Guitar A	Guitar B	Guitar C
Guitar A	**213**	4	89
Guitar B	243	**2**	79
Guitar C	242	0	**82**

The nontonal peak classificator tends to classify everything as guitar A.

Set 3: f_0, MFCC, 501 Features, Performance = 75.3%

	Guitar A	Guitar B	Guitar C
Guitar A	**223**	55	46
Guitar B	52	**229**	43
Guitar C	28	16	**280**

The MFCC classificator is the single-feature classificator with the highest rate of correct classifications.

Set 4: f_0, NT, PA, 671 Features, Performance = 57.3%

	Guitar A	Guitar B	Guitar C
Guitar A	**173**	83	68
Guitar B	105	**150**	69
Guitar C	49	41	**234**

Compared to set 1, the identification of guitar B has improved.

Set 5: f_0, MFCC, PA, 1141 Features, Performance = 81.6%

	Guitar A	Guitar B	Guitar C
Guitar A	**241**	43	40
Guitar B	32	**265**	27
Guitar C	17	20	**287**

This is the classificator with the best overall performance.

Set 6: f_0, MFCC, NT, 531 Features, Performance = 77.6%

	Guitar A	Guitar B	Guitar C
Guitar A	**238**	50	36
Guitar B	46	**235**	43
Guitar C	24	19	**281**

The NT features slightly improve the performance of the MFCCs, especially for distinguishing guitar A from guitar B.

Set 7: f_0, MFCC, PA, NT, 1171 Features, Performance = 80.6%

	Guitar A	Guitar B	Guitar C
Guitar A	**248**	38	38
Guitar B	38	**257**	29
Guitar C	21	25	**278**

Even though the number of features is higher as in set 5, the overall performance is slightly less.

Human Listeners, Performance = 65%

	Guitar A	Guitar B	Guitar C
Guitar A	**4**	6	2
Guitar B	2	**10**	0
Guitar C	1	1	**10**

Human listeners tend to mistake guitar A for guitar B.

An analysis of the misclassified samples showed no correlation between player, string, fret, or timbre and the misclassification rate.

As already mentioned, the classification is based on one single note, without knowledge of timbre, string or fret position. In a realistic setup the classification will be based on *several* notes on different strings and with different timbres.

With a single-tone misclassification probability p_1 of 0.2 (from sets 5 or 7), a classification based on *three* tones with a 2-out-of-3 selection would lead to a misclassification probability p_3 of

$$p_3 = (1 - p_1)^3 + 3 \cdot p_1(1 - p_1)^2 = 0.10 \quad (1)$$

thus the classification performance based on three tones will be 90%!

5. CONCLUSIONS

It is possible to distinguish individual classical guitars by single sound samples. The identification can be accomplished by Support Vector Machines executing a 1-vs-1-classification using feature vectors with time-spectral properties of guitar sounds.

For an identification based on single notes a classification performance of better than 80% is obtained, far outperforming the results of human classification (65%).

A next task will be the extension of the database to more guitars. It will have to be investigated whether reliable SVMs can be created based on the recordings of fewer guitarists, possibly only one. Also the robustness of the classification against different recording setups will have to be tested, because it is highly desirable to record the unknown sounds in a normal living room or a workshop.

After this, the system is ready to be tested in a luthier's workshop. The instrument under construction will be classified using a SVM which was trained with the data of a reference instrument. The luthier will successively modify his instrument until it will be classified as the reference instrument.

6. ACKNOWLEDGEMENTS

We kindly thank the luthier Michael Wichmann, Hamburg for providing us with the test instruments and for arranging the contact to the artists. We thank the artists Annika Hoffmann, Satoshi Oba, Heiko Ossig, Bernhard Schleiser, and Christian Schulz for recording the sound samples.

7. REFERENCES

[1] C. Traube and J. O. Smith, "Extracting the fingering and the plucking points on a guitar string from a recording," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01)*, 2001.

[2] Rolf Bader, "Physical model of a complete classical guitar body," in *Proc. of the Stockholm Music Acoustics Conference, August 6-9, 2003 (SMAC 03), Stockholm, Sweden*, 2003, pp. 121–124.

[3] Bernard E. Richardson, "Guitar models for makers," *Proc. of the Stockholm Music Acoustics Conference, August 6-9, 2003 (SMAC 03), Stockholm, Sweden*, pp. 117–120, 2003.

[4] T. Ono and D. Isomura, "Acoustic characteristics of carbon fiber-reinforced synthetic wood for musical instrument soundboards," *Acoustical Science and Technology*, vol. 25, no. 6, pp. 475–477, 2004.

[5] Cumhur Erkut, Vesa Välimäki, Matti Karjalainen, and Mikael Laurson, "Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar," in *AES108th Convention, Paris*, 2000.

[6] M. Karjalainen, H. Penttinen, and V. Valimaki, "Acoustic sound from the electric guitar using DSP techniques," *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE Int. Conference on*, vol. 2, 2000.

[7] D. van Steelant, K. Tanghe, S. Degroeve, B. De Baets, M. Leman, and JP Martens, "Classification of percussive sounds using support vector machines," *Proc. of the Annual Machine Learning Conference of Belgium and The Netherlands, Brussels, Belgium*, 2004.

[8] Janet Marques and Pedro J. Moreno, "A study of musical instrument classification using gaussian mixture models and support vector machines," Tech. Rep., Cambridge Research Laboratory, 1999.

[9] D. Deng, C. Simmermacher, and S. Cranefield, "Finding the right features for instrument classification of classical music," *Proc. of the IEEE Int. Workshop on Integrating AI and Data Mining AIDM'06*, 2006.

[10] 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*, vol. 14, no. 3, 2006.

[11] Malcolm Slaney, "Auditory toolbox. a MATLAB toolbox for auditory modeling work. version 2," 1998, URL: http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/, visited June 2007.

[12] Bernhard Schölkopf and Alexander J. Smola, *Learning with Kernels. Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press, 2002.

[13] Thorsten Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds. MIT-Press, 1999, URL: http://svmlight.joachims.org/.

[14] J. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAGs for multiclass classification," in *Advances in Neural Information Processing Systems 12*, S.A. Solla, T.K. Leen, and K.-R. Mueller, Eds., 2000, pp. 547–553.