Automatic Recognition of Tonal Instruments in Polyphonic Music from different Cultural Backgrounds

Matthias Kahl, Jakob Abeßer, Christian Dittmar, Holger Großmann

Fraunhofer Institut für Digitale Medientechnologie, 98693 Ilmenau, Germany, Email: {kahlms,abr,dmr,grn}@idmt.fraunhofer.de

Abstract

This paper describes a feature-based approach for the note-wise classification of instruments in a polyphonic music segment. To reduce the amount of potential misclassification due to spectral overlapping of different instruments, we perform a pre-selection of notes. In doing so, we allow at most two instrument notes to overlap at the same time. If a dominant instrument is present, we try to detect it according to a majority decision of all note-wise classification results. Two evaluation experiments for selections of both western and non-western instruments are performed to examine the applicability of the implemented features for instrument recognition in a world-music context. For this purpose, we used three publicly available large-scale databases of instrument sounds. They contain samples recorded with varying instruments, recording conditions, performers, dynamic range and playing techniques. Furthermore we compile a novel database for non-western instruments. In addition to well-established audio features such as *MPEG-7* features and *MFCC*'s, we introduce a selection of novel features. For instance, the course of the envelopes is characterized for each harmonic separately and in comparison to the fundamental frequency.

Motivation

Information about the Instrumentation of a music segments can be used as high level feature for genre classification. In addition to that, the query for certain instrumental parts (e. g. guitar solo) is a desirable functionality in media player software. In this paper, we describe an approach for the note-wise classification of instruments in a polyphonic music segment.

Preprocessing

Because of the use of harmonic features, the stable fundamental frequency Algorithm *FFEMM* [3] was used. This Algorithm is based on maximizing the mean of the fundamental frequency hypothesis and the pitch dependent count of harmonics in a median filtered spectral envelope. *FFEMM* gave the best results in this scenario compared to other algorithms [3].

To evaluate whole notes inside music segments, we used an onset detection based on linear prediction [2]. One of the biggest challenges for the detection of instruments in polyphonic and multitimbral music segments is the spectral influence from the accompaniment instruments. Thus, we detect the fundamental frequency of the considered note and muted all frequencies outside its fundamental frequency and harmonics with an comb filter.

Acoustic Features

In addition to well described features like *MPEG-7* Audio Descriptors [1] and *MFCC*'s, we designed novel descriptors focusing on harmonic characteristics of single notes.

Some string instruments show some slight frequency amplitude modulations in their harmonics during the decay part of the note. We computed a flatness measure of the energy slope in each harmonic to determine the intensity of this modulation.

Every instrument shows a characteristic distribution of energy across its harmonics. Thus, we take the ratio of each harmonic energy to the fundamental frequency energy.

As an additional feature, we check for the existence and intensity of the tremolo effect. Therefore we conducted a frequency analysis on the course of the fundamental frequency over time within a range of $2Hz$ to $12Hz$.

Many wind instruments exhibit a noticeable breath noise. We investigate the dominance of this noise by calculating the ratio between the energy in the tonal parts to the energy in the noise parts.

Several other spectral, temporal and harmonic features like *Spectral Flux*, *Zero Crossing Rate*, *Inharmonicity*, *Even-Odd Harmonics* etc. were used as well to assemble a feature set of 260 coefficients in total.

Classification

To reduce the calculation time and dimansionality of the feature space, we perform a Linear Discriminant Analysis. Therefore, the system is reduced to n-1 dimensions where n denotes the number of classes. To evaluate the system we performed a classification with a *Support Vector Machine* (SVM) and the *Learning Vector Quantization* (LVQ) in a accuracy cross validation. In all performed experiments, the *SVM* classifier outperformed the LVQ thus we only used the svm for evaluation.

System

The aim of our approach is to detect the main instrument of the analysed music segment. In this case, the main instrument exhibits the most onsets and the highest dynamic of the tonal parts in this music segment.

Because of the assumption that one music part is often consistently played by one single instrument, we took a majority decision.

To evaluate the detection in the case of isolated notes, we used a combination of the RWC database [5], the samples of the University of Iowa [4] and the Master Samples of the McGill University [6]. To assess the separation quality in the case of polyphonic and multitimbral music segments, we created a set of mixtures from two isolated notes. The accompanying instrument is reduced in its amplitude to distinguish the main and accompanying instrument. In addition to the database combination, we used an internal and annotated real music database. We used the six instruments FLUTE, GUITAR, PIANO, SAXOPHONE, TRUMPET and VIOLIN in the first class taxonomy. In a further experiment we used the taxonomy of the North African OUD, the Japanese SHAMISEN and SHAKUHACHI, the Indian SITAR and the European CLARINET and ACOUSTIC GUITAR.

Experiments

We performed an evaluation about all three cases and the two taxonomies. for the evaluation using isolated notes, we achieved best classification accuracy of 92.7 %. In case of artificial mixed sounds, we achieved 77.38 % and in case of using notes from real music, we achieved 83.33 % for using cross validation over all samples and 53.87 % for strictly separating songs for training and test purpose. with notes of real music we achieved 83.33 with cross validation and 53.87 % with unknown soloparts. With use of the majority decision, we achieved 72.40 % about the unknown soloparts. These results were given by the evaluation process on the taxonomy 1. The summary is shown in Table 1.

train-\ testset	isolated	art.mix	real music
isolated	92.28 %	58.76 %	20.70 %
art.mix	87.73 %	77.38 %	42.20 %
real music	27.83 %	35.13 %	52.99 %
combination	90.26 %	73.46 %	50.05 %

Tabelle 1: comparison of different train- and testsets

The results show that the classifier gave the best results, when the trainset and testset of the same kind. The combination of all three training data sets lead to the most flexible classifier.

We used the same system to classify the instruments from the second taxonomy. In this case we made a one track cross validation, which means that each fold corresponds to one music segment. This is necessary to prevent that some notes of the music segment are in the trainset and some other notes from the same music segment in the testset. For this experiment we achieved 79.72 %.

Conclusion

In this work, we presented a system to detect the main instrument in polyphonic and multitimbral music segments. On isolated notes, we achieved a classification accuracy of 92.70 %, on artificial mixed sounds 77.38 %, on notes from real soloparts 83.33 %, on notes from unknown soloparts 53.87 % and with use of majority decision 72.40 %. In the world music instrument taxonomy, we achieved a accuracy result of 79.72 % and with use of majority decision 92.05 %. To improve the system, a finer taxonomy should be applied to cope with different playing techniques or instrument manufactures. A better note extraction, the use of high level features such as chord or single note playing and a source separation can improve the classification as well.

	cl	*gt*	*ou*	*sk*	*sm*	*st*
cl	75	5	4	12	1	4
gt	3	81	11	1	0	4
ou	4	14	80	1	0	1
sk	10	2	1	79	2	5
sm	0	0	0	15	84	1
st	5	8	1	9	0	77

Tabelle 2: confusion matrix for the second taxonomy in % with *cl* (clarinett), *gt* (guitar), *ou* (oud), *sk* (shakuhachi), *sm* (shamisen), *st* (sitar). horizontal: ground truth, vertical: classified instruments

Acknowledgements

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). Additionally, 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, enabling transnational cooperation between Thuringian companies and their partners from other European regions.

Literatur

[1] Martinez, J. M.: *MPEG-7* Overview, October 2004. `http://chiariglione.org/mpeg/standards/mpeg7/mpeg7.htm`

[2] W.Chi Lee und C.-C. J. Kuo, Musical Onset Detection based on Adaptive Linear Prediction, ICME, 2006.

[3] Kahl, M.: Klassifikation von Hauptinstrumenten in polyphonen und multitmbralen Musikaufnahmen 2010. thesis, Fraunhofer IDMT, FH Schmalkalden

[4] Fritts L.: `http://theremin.music.uiowa.edu/mis.html` 1997. (last viewed: 18.03.2010)

[5] Goto M. et. al.: RWC Music Database: Music Genre Database and musical Instrument sound Database. Proceedings of the *ISMIR*, 2003.

[6] Opolko, F. und Wapnick, J.: McGill University Master Samples, McGill University 1987.