Instrument classification using Hidden Markov Models

Matthias Eichner, Matthias Wolff, Rüdiger Hoffmann
Technische Universität Dresden
Laboratory of Acoustics and Speech Communication
{eichner,wolff,hoffmann}@ias.et.tu-dresden.de

Abstract

In this paper we present first results on musical instrument classification using an HMM based recognizer. The final goal of our work is to automatically evaluate instruments and to classify them according to their characteristics. The first step in this direction was to train a system that is able to recognize a particular instrument among others of the same kind (e.g. guitars). The recognition is based on solo music pieces played on the instrument under various conditions. For this purpose a database was designed and is currently being recorded that comprises four instrument types: classical guitar, violin, trumpet and clarinet. We briefly describe the classifier and give first experimental results on the classification of acoustic guitars.

Keywords: Automatic musical instrument recognition, music content processing, multimedia content description

1. Introduction

Robust musical instrument recognition could lead to a variety of applications in the field of music content analysis including automatic annotation of musical signals, retrieval of music from a database or quality assessment of instruments. Different approaches have been investigated in the past that differ in the features used to describe the important spectral and temporal properties of the signal, the classification strategy and the musical pieces that were used to train the system [1, 2, 3].

In this paper we apply former work on a general structure discovering technique for speech signals [4] to music signals. The method infers Hidden Markov Models (HMMs) of arbitrary topology in an entirely data-driven way from a set of training signals. This is particularly useful if there is few or no prior knowledge about the temporal structure of the signals to model. We successfully applied this technique to supersonic signals used in process integrated non-destructive testing (PINT) tasks [5].

The paper is organized as follows. Section two describes the database used for the experiments. The classifier is briefly described in section three. Finally, first experimental results on note recognition and on the identification of guitars are presented and discussed.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.
© 2006 University of Victoria

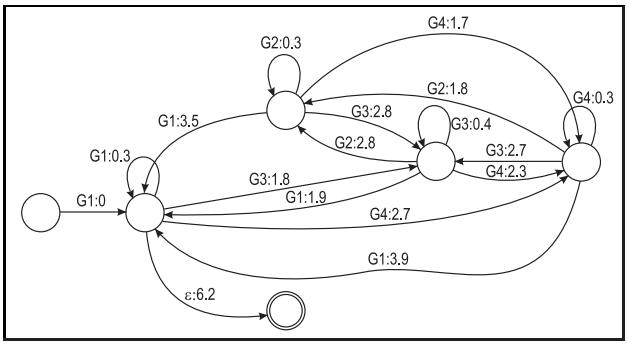

Figure 1. HMM automaton graph for one individual guitar. The transitions are annotated with the index of a Gaussian PDF defining one state in the feature space and a weight (neg. log. transition probability).

2. Database

For our experiments we designed a database that comprises four instrument types: classical guitar, violin, trumpet and clarinet. Those instruments were chosen because (a) they cover a broad range of instrument groups and (b) we have several instruments permanently available for each group. This allows us to add new recordings in the future. For every instrument type we do 600 recordings varying the following conditions: (a) 10 instruments, (b) 2 rooms (unechoic and conference room), (c) 3 solo pieces, (d) 5 players and (e) 2 repetitive playings.

So far, we finished the recordings for the classical guitars and are currently working on the violins. The selected guitars cover a wide spectrum in construction type and sound characteristics. We used an artificial head which was placed 2 meters in front of the musician for the recordings. There were three solo pieces (a scale, a blues and an etude) played on the guitars. The pieces are all approximately 30 s long and contain only few polyphonic parts. We labeled all recordings on note level in a semi-automatic procedure.

3. Training and Classification

We use our standard speech recognizer for the experiments and trained acoustic models for (1) single notes and (2) individual instruments. First we pass the recordings through a 31 channel mel–scaled filter bank (which slightly outperforms MFCCs in our speech recognizer) and compute the

Table 1. Experiment 1 - Note recognition correctness depending on the number of Gaussian PDFs.

# GMs	Correctness	Density
532	78%	2.97
1061	76%	3.17

first and second order differences which results in a 63-dimensional primary feature vector. Then we apply a statistical principal component analysis and reduce the feature space to 25 dimensions.

We tested two different sets of HMMs. For a first experiment we trained HMMs for all notes occurring in the recordings independently of the instrument and all other conditions described in section 2. The recognition task was to find the most likely sequences of notes for unseen recordings using the Viterbi algorithm. The correctness of the result was assessed by the standard DTW string alignment to semi-automatically generated reference labels.

In a second experiment we trained one HMM for each instrument independently of all other conditions. Here the recognition task was to identify one out of the ten instruments of one type by an unseen recording. This was done computing the most likely state sequence of the recording in all instrument HMMs and selecting the one with the highest likelihood score.

In both experiments we applied an HMM inference procedure which is able to discover the structure of signals and – in contrast to [3] – to model it symbolically by inferring not only the Gaussian PDFs and the transition matrix of the HMMs by also an arbitrary automaton graph [5]. Figure 1 shows a very simple example for an HMM topology modeling a particular guitar (second experiment).

4. Experiments

The experiments discussed in this section were carried out using the 600 recordings of guitar solo pieces as described in section 2.

For the first experiment we used the 120 recordings of one guitar player. We randomly picked 110 recordings as training set and 10 as test set and trained 177 note HMMs and one silence model. Table 1 shows the note recognition correctness for two different sets of HMMs. Despite the high label insertion rate of about 300 % the experiment prooved that our recognizer is suitable for processing musical signals.

In the second experiment we trained one HMM per guitar and one silence model. The task was to find out on which of the ten guitars test recordings not seen by the training were played. Table 2 shows the guitar identification results for models trained with data from either a single or two musicians. The classification margin was calculated by averaging the differences between best and the second-best score

Table 2. Experiment 2 - Guitar identification correctness depending on the number of Gaussian PDFs per guitar model and the number of musicians in training set.

	Single Musician		Two Musicians	
# GMs	Correct	Margin	Correct	Margin
15	60%	0.54	65%	0.26
30	70%	0.82	75%	0.34
60	100%	1.02	85%	0.48
120	100%	1.40	95%	0.62
238	100%	1.66	85%	0.69

of all correct classified guitars. For recordings of a single musician the system is able to correctly indentify all instruments. If recordings of another musician are added to the training set the performance decreases.

5. Conclusion

The experiments described are preliminary and work in progress, but they show the suitability of the chosen approach. Further research will include experiments using more data and building better models for instrument recognition using label information. We will also try to develop a strategy to evaluate unseen instruments and to assign them descriptive attributes. The used dataset has a very strong influence on the obtained results. Therefore, we plan to verify our experiments to publicly accessible databases.

6. Acknowledgments

This research is supported by the BMBF grant 03i4745A and is conducted in cooperation with the Institut für Musikinstrumentenbau Zwota, Germany (IfM).

References

[1] P. Herrera-Boyer, X. Amatriain, E. Batlle and X. Serra, “Towards Instrument Segmentation for Music Content Description: a Critical Review of Instrument Classification Techniques”, 1st International Conference on Music Information Retrieval, 2000.

[2] A.G. Krishna and T.V. Sreenivas, “Music instrument recognition: from isolated notes to solo phrases”, Proceedings ICASSP-04 (IEEE International Conference on Acoustics, Speech and Signal Processing), 2004.

[3] M. Casey, “Generalized Sound Classification and Similarity in MPEG-7”, Organized Sound, 6:2, Cambridge University Press, 2002.

[4] M. Eichner, M. Wolff and R. Hoffmann, “A unified approach for speech synthesis and speech recognition using stochastic Markov graphs,” in Proc. 6th Int. Conf. Spoken Language Processing (ICSLP), vol. 1, pp. 701–704, Beijing, PR China, 2000.

[5] C. Tschöpe, D. Hentschel, M. Wolff, M. Eichner and R. Hoffmann, “Classification of non-speech acoustic signals using structure models”, Proc. IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP), 2004.