

# Automated musical instrument recognition

Alireza Amiri

Georg-August-University Göttingen  
Institute of Computer Science  
Göttingen, Germany  
alireza.amiri@stud.uni-goettingen.de

Thorben Janz

Georg-August-University Göttingen  
Institute of Computer Science  
Göttingen, Germany  
t.janz@stud.uni-goettingen.de

Dimitra Despoina Maoutsa

Georg-August-University Göttingen  
Institute of Computer Science  
Göttingen, Germany  
d.maoutsa@stud.uni-goettingen.de

## I. INTRODUCTION

The expanding volume of available musical data leads to a rapid growth of the research field of automated music information retrieval due to the need of efficient administration and exploration of musical content. This need results from the change of the way people listen to and interact with music. Modern music streaming applications enable the user to listen to music, which is selected based on tags the user indicates.

One important aspect of music information retrieval is the automated musical instrument recognition. It has several applications, such as automatic musical transcription. This process consumes a lot of effort and time for human beings. Therefore a tool supporting the extraction of musical scores is capable of saving time thus money.

The project discussed in this paper deals with the automated recognition of musical instruments. The goal is to hierarchically classify musical instruments according to their family (strings, woodwinds etc.) and furthermore, in case of success, to proceed to intra-family instrument distinction. This will be achieved by developing a classifier which uses the machine learning approaches. This project is part of the master's level class "Machine Learning and Pervasive Computing" at the Georg-August-University of Göttingen.

Obviously audio signals are needed as the input data. One alternative is recording audios by microphone. Another choice is using previously recorded audio files. In this project the second option is used due to the lack of high-quality devices. Therefore audio files represent the more suitable choice, to reduce misclassification caused by noisy data. Consequently no sensors are required to obtain the input data.

The majority of the research projects on this field is based either on the data of single note recordings or isolated solo-performances of single instruments, which are referred to as "monophonic" recordings. Only a small part of researches deals with polyphonic recordings, where two or more instruments are played simultaneously. In terms of recognizing instruments from monophonic recordings, most problems are solved accurately. Accordingly the present focus is set on polyphonic instrument recognition. In addition

polyphonic recordings are closer to real-world music. For that reason polyphonic audio files are in this project's interest.

The novelty of our approach is the additional use of context information besides content information. Namely we are going to include information regarding the musical genre in order to provide some prior knowledge about the possible instrumentation of the songs of the recordings.

## II. RELATED DOCUMENTS

The scientific work on automated musical instrument recognition has its roots in the late 70s. In this period of time, the basic methods and techniques for this classification problem were developed. The first attempts tried to map the sounds of musical instruments to a three-dimensional space, using subjective similarity judgments (Grey, 1976).

The research field had its height during the late 90s, when the existing methods were applied for processing mainly monophonic audio signals. Based on the early works, the effort on recognizing instruments from single note recordings, respectively monophonic real-world recordings, was made (De Poli, et al., 1993). Additionally, compared to the earlier attempts, the researches and projects used machine learning approaches to train their classifier.

Observations show that the accuracy of classifiers developed for monophonic signals, is significantly lower when applied on polyphonic recordings. In (Kitahara, et al., 2005) three problems are stated, which need to be solved to achieve a high success rate in classification:

- feature variations
- timbre's dependency on the pitch
- musical context

The variation of features is the main problem when processing polyphonic audio signals. It is caused by the overlapping of and interference between the frequencies of the single instruments. Therefore it is not possible to separate the sounds of every single recorded instrument without distorting the acoustic signals. If this would be possible, the problem of polyphonic signal processing could be reduced to monophonic signal processing.

The approach to solve this problem which is mentioned in (Kitahara, et al., 2005) gives every feature a weight. Features that are more likely to have high variance within a class due to overlapping frequencies get a lower weight. This is achieved by applying linear discriminant analysis.

The second issue, the timbre's dependency on the pitch, results from the wide pitch range of many instruments over many octaves. Therefore a pitch-dependent timbre model should be used to deal with this problem.

The consideration of neighboring notes when classifying the instrument playing a single note of a melody is a possible solution for the third problem. When identifying single notes of a melody, it can happen that the instrument playing one of these notes is classified differently as the others. Since this is very unlikely in musical way, a probability function should be taken into account, to decide if a likewise classification result is correct.

The paper (Bosch, et al., 2012) compares different techniques to separate different overlapping frequencies. This leads to the previous mentioned issue of feature variations. Consequently the more precise the technique separates the signals the lesser is this issue's impact. As it concludes, the FASST algorithm fits for separating the audio signals when attempting to recognize the predominant pitched instrument. It separates the input audio into the four parts:

- drums
- bass
- melody
- remaining sounds

The melody-part contains the signal of the predominant instrument separately. Accordingly this signal can be used as the input data for the classifier.

In (Okamura, et al., 2012) Mari Okamura proposed a novel approach called "Example-based Sparse Representation" in which source separation is not applied. This approach uses sample feature vectors of different musical instruments as the base matrix of sparse representation. The accuracy of 91.9% was obtained for monophonic sounds. Also in polyphonic sounds 51.1% of combinations were estimated correctly.

When attempting to differentiate between larger numbers of instruments, a used approach in research is the hierarchical classification approach. As described in (Eronen & Klapuri, 2000) it uses several levels of abstraction, to particularize the classified instrument step by step. On one of the first levels, the decision may be based on the musical instrument family, while low-level decisions distinguish between the single instruments themselves.

Applying this approach, it is possible to construct subsets of features that are used on different nodes. Thus only well-suited features for the specialized decision in every node are considered by the classifier.

The crucial part applying this approach is the appropriate selecting of the features for deciding single steps. In addition,

the way the different levels are built is important to achieve good results.

The usefulness of the hierarchical approach is proven by many projects in the past.

### III. REFERENCES

- Bosch, J., Janer, J., Fuhrmann, F. & Herrera, P., 2012. *A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals*, Barcelona: Music Technology Group, Universitat Pompeu Fabra.
- De Poli, G., Prandoni, P. & Tonella, P., 1993. *Timbre clustering by self-organizing neural networks*, Milan: University of Milan.
- Eronen, A. & Klapuri, A., 2000. *Musical instrument recognition using cepstral coefficients and temporal features*, Tampere: Signal Processing Laboratory, Tampere University of Technology.
- Grey, J. M., 1976. *Multidimensional perceptual scaling of musical timbres*, Stanford: Center for Computer Research in Music and Acoustics, Department of Music, Stanford University.
- Kitahara, T. et al., 2005. *INSTRUMENT IDENTIFICATION IN POLYPHONIC MUSIC: FEATURE WEIGHTING WITH MIXED SOUNDS, PITCH-DEPENDENT TIMBRE MODELING, AND USE OF MUSICAL CONTEXT*, Kyoto: Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University.
- Okamura, M., Takehara, M., Tamura, S. & Hayamizu, S., 2012. *Toward polyphonic musical instrument identification using example-based sparse representation*, Gifu: Department of Information Science, Gifu University.