Musical Instrument Recognition in Polyphonic Music

Toni Heittola

Department of Signal Processing
Tampere University of Technology
P.O. Box 553, FI-33101 Tampere, Finland
e-mail: toni.heittola@tut.fi

1 Introduction

Understanding the timbre and pitch of musical instruments is an important issue for automatic music transcription, music information retrieval and computational auditory scene analysis. In particular, recent worldwide popularization of online music distribution services and portable digital music players makes musical instrument recognition even more important. Musical instruments are one of the main criteria (besides musical genre), which can be used to search certain type of music from music databases. Some classical music are even characterized with the used musical instruments (e.g. piano sonata and string quartet).

There has been only very little research to determine the human ability to discriminate between the sounds of different musical instruments. Martin *et al.* conducted two listening experiments in [3]. Test subjects were asked to classify isolated notes (monotimbral and monophonic sounds) from fourteen instruments into 28 possible instrument classes. The overall recognition accuracy for individual instruments were 46% and for instrument families 92%. With 10-second long excerpts with multiple notes recognition accuracies increased to 67% and 97%.

Most of the research on automatic recognition of musical instruments until now has been carried out using isolated notes. However, in recent years there has been increasing amount of research dealing with instrument-labeling in more complex music signals, such as monotimbral phrases, duets, or even realistic music signals [2].

The purpose of my research is to develop mathematical models for sound sources and apply these in the automatic analysis and coding of polyphonic music. Target signal are musical signals and in limited cases also speech signals. The redundant frequency information of the harmonic sounds will be used in the developed new models. The developed modeling schemes will be tested in two applications, musical instrument recognition in polyphonic music and in music transcription.

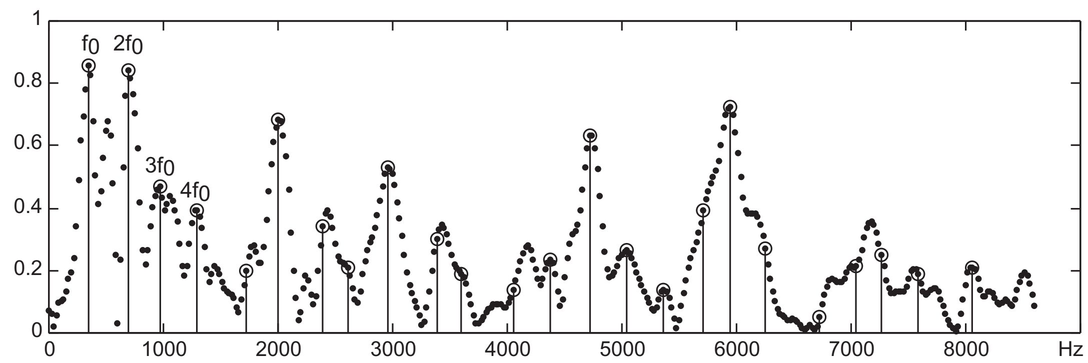

Figure 1: FFT of the note sample (piano) with fundamental frequency (F0) 332 Hz. Picked spectral peaks are marked with circle in the figure.

2 Proposed Approach

Redundant frequency information of the harmonic sounds was used as starting point in the development of robust features for the instrument recognition in the polyphonic case. Based on the estimated fundamental frequency maximum spectral peaks are picked within 150 Hz search area around the theoretical harmonic frequencies (see Figure 1). Such a wide search area is used to allow inharmonicity of certain instruments (e.g. piano and guitar). Finally, Mel-Frequency Cepstral Coefficients (MFCC) and their time-derivatives are calculated for the note.

With certain note combinations, there will be shared frequency components between overlapping notes. Thus two more features will be extracted, one where every second harmonic frequency component is leaved out and one where every third harmonic frequency component is leaved out while harmonic spectral peaks are collected. This will capture at least some information about the individual harmonic components.

Gaussian Mixture Models (GMM) are used to model instrument-conditional densities for these features. In the classification stage, likelihoods of the three features for the particular note are combined and classification is done with maximum-likelihood classifier.

3 Experiments

Ten instrument classes (piano, electric piano, guitar, electric guitar, electric bass, violin, cello, saxophone, oboe, flute) were selected for the experiment. Thousand note combinations were randomly generated allowing all possible note and instrument combinations (excluding unison). Instrument samples were randomly selected from joined database of the commercial McGill University Master Samples collection (MUMS) [4], University of Iowa sample collection, IRCAM's Studio On-line, and Real World

Computing (RWC) database [1]. The database contains 63 individual instrument instances and 579 played note scales.

In order to maximize training data, classification was carried out using leave-one-out method. Instrument instances used in the note combination sample to be classified were excluded from the training data. Only 200ms long excerpts from each note sample were found to give adequate results. In the classification stage, fundamental frequencies of the individual notes in the note combination was known.

The system performed quite well for the two simultaneous notes, recognizing 41 % of the notes. The performance for three simultaneous notes was slighly lower 32 %. In the future, we try to extend the system for higher polyphony. This requires an intelligent feature selection, i.e. selecting when to use all or reduced amount of harmonics in the feature extraction. Since music mainly consist of consecutive notes played with same intruments, this valuable information has to be taken account also in the recognition process.

References

[1] M. Goto, T. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Music genre database and musical instrument sound database. pages 229–230, Baltimore, USA, October 2003. International Conference on Music Information Retrieval (ISMIR).

[2] Perfecto Herrera-Boyer, Anssi Klapuri, and Manuel Davy. Automatic classification of pitched musical instrument sounds. In *Signal Processing Methods for Music Transcription*, pages 163–200. Springer.

[3] K. Martin. *Sound-source recognition: A theory and computational model*. PhD thesis, Massachusetts Institute of Technology, 1999.

[4] F. Opolko and J. Wapnick. Mcgill university master samples. Technical report, McGill University, Montreal, Canada, 1987.