Instrument Identification

MUS-15

Ioanna Karydi

July 10, 2015

Introduction

This report presents an overview on musical instrument identification techniques. In the beginning, it discusses the basic characteristics and parameters that are building blocks of a sound. Later on, we attempt to present how well humans are capable of performing the task of instrument identification. We would then examine the characteristics that are needed for a perfect algorithm to perform this task.
These days, there are multiple approaches and techniques that aim to tackle the challenge of instrument identification. We have picked two techniques and towards the end of this report, we would present an overview for these techniques. This would give a clear picture of how these techniques work, the steps that they follow until successful identification and their respective performances.

Sound attributes and Timbre importance

In order to be able to identify a sound, we need to use characteristics that describe it. There are four basic attributes that are used for this purpose.

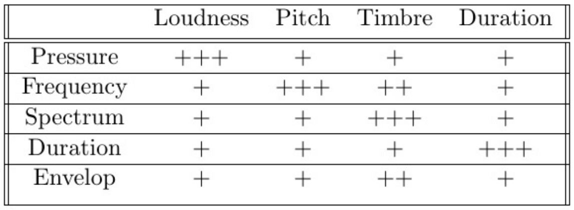
Loudness: "That attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud".(American National Standards Institute, "American national psychoacoustical terminology" S3.20, 1973, American Standards Association.)

Pitch: Related to the frequency of the sound, pitch can characterize a sound from high to low. It can order the sounds in a scale related to frequencies.

Duration: The attribute that describes how long or short the notes last. It can also be used to describe how long the whole piece of music lasts. However, it generally describes the length in sense of time a sound lasts.

Timbre: Timbre is the most complex quality of a sound. American Standards Association defines timber as: "Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar."

The following table shows the connection between sound attributes and measured physical parameters:



	Loudness	Pitch	Timbre	Duration
Pressure	+++	+	+	+
Frequency	+	+++	++	+
Spectrum	+	+	+++	+
Duration	+	+	+	+++
Envelop	+	+	++	+

Figure 1: Attributes of sound dependence on physical parameters

From the above table, it is obvious that usually there is only one parameter that affects each attribute (except for timbre). That helps us perceive the meaning of each one of them more easily. On the other hand, timbre is a multidimensional attribute of sound that makes the way humans perceive it more difficult.

Human performance

Musical instrument identification systems attempt to reproduce how humans can recognize and identify the musical sounds populating their environment[1]. Hence examining the performance of humans can be seen as good way of evaluating success rate of the existing systems. A research performed using isolated sounds from several orchestral musical instruments calculated the performance of 88 experienced listeners. As expected, the identification rate was dramatically decreasing when the number of musical instruments was increasing. For example a set of 27 instruments was used to test the performance, the success rate was 55.5%. However, using a set of less than ten musical instrument improved the identification rate dramatically, exceeding 90%.

The ideal Algorithm

It is generally hard to compare different systems used for instrument identification, since investigators evaluate their system with the data sets at their disposal. In addition there are different methodologies used for the evaluation procedures. If we could describe an ideal algorithm, among the generally acceptable competences, it should have the following characteristics:

Generalization, refers to the stable performance of the algorithm, given a system trained with a specific dataset, even when using subsequent unknown data.

Robustness, this characteristic has to do with performance of the algorithm. It recognizes the same sound under different recording conditions including pitch, quality, playing style etc.

Meaningful behavior, the algorithm is desirable to behave in a meaningful way. This means it should be as close as possible to the way humans perform in the task of recognition. A good example would be the confusion between instruments or instrument families.

Reasonable computational requirements, in case of including the algorithm into wider Musical Instrument Recognition (MRI) frameworks, the responsiveness or the latency of overall structure should not be significantly affected.

Modularity, represents the ease and flexibility to update the system with new samples at any given time. Likewise, no retraining should be required in case of new instrument additions to the system.

Selected Approaches

Two approaches have been chosen to be presented in this report. Each one of them for different reasons. The first approach from G. Agostini, M. Longari, and E. Pollastri, uses monophonic sounds and has an interestingly high performance. The second one from Jayme Garcia Arnal Barbedo and George Tzanetakis, recognizes polyphonic sounds and it presents an interesting approach that is not using the traditional machine learning methods, introducing a new promising technique.

G. Agostini, M. Longari, and E. Pollastri Approach

This approach features that are extracted, describe spectral characteristics of monophonic sounds. Monophony is a melody without harmony. This may be realized as one note at a time, or the same note played at different octaves. The dataset that is used in this approach is composed of 1007 tones using 27 musical instruments. A wide range of sounds were used, ranging from orchestral sounds (strings, woodwinds, brass) to pop/electronic instruments (bass, electric and distorted guitar). The classification of features was conducted using widely used pattern recognition techniques like Discriminant Analysis (Canonical discriminant analysis and Quadratic Discriminant Analysis), Support Vector Machines and k-nearest neighbors. Each one of these techniques showed different performance results with Quadratic Discriminant Analysis having least error rates.

Feature extraction

A set of features related to the harmonic properties of sounds is extracted from monophonic musical signals. This is divided in three stages a brief description, each one of them is as follows:

- **Audio Segmentation**

  This stage focusing on the temporal segmentation of the signal into a sequence of meaningful events. This is achieved by: cutting-off the silences that might appear and cutting-off frequencies to filter out unwanted noise that might come from vibrations. In case of failure to detect some tone transitions, the next stage can take care of this.

- **Pitch Tracking**

  Pitch tracking helps to refine the previous step. Pitch is a basic attribute that is used to calculate some spectral features. This steps results in three outcomes: calculation of the average value for each note hypothesis, a value of pitch for each frame and an accuracy value that represents the uncertainty of estimation.

- **Calculation of Features**

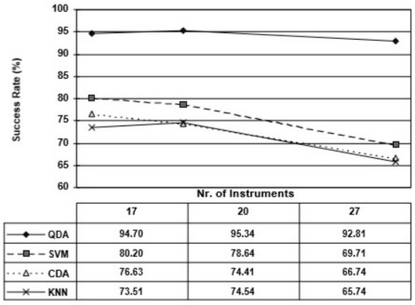
  After completing the previous steps, the tones are isolated and for each one of them a set of 9 features is extracted. Their mean and standard deviations for the purpose of describing this events is also calculated. This results in a total of 18 features for each tone. For analyzing the signal half-overlapping windows are used and Hamming functions used for smoothing it. Finally, a Short-Time Fourier Analysis is applied for spectrum estimation.

Classification Techniques

For classifying the data, four of the most popular classification techniques were used: k-Nearest Neighbors, Support Vector Machines, Canonical Discriminant Analysis and Quadratic Discriminant analysis. Each one of these resulted in different success rates that is presented in the following paragraph.

Performance

The performance for each one of the four experiments was evaluated, using datasets of 17, 20, 27 instruments. QDA (Quadratic Discriminant Analysis) performed better in every test compared to the other classifiers. The success rate of this classifier was really high, even for 27 instruments, reaching 92.81%. The best results for each one of the classifiers are shown in the following figure:



	17	20	27
QDA	94.70	95.34	92.81
SVM	80.20	78.64	69.71
CDA	76.63	74.41	66.74
KNN	73.51	74.54	65.74

Figure 2: Classifiers performances for different number of instruments

Jayme Garcia Arnal Barbedo and George Tzanetakis Approach

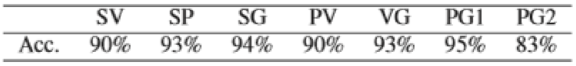
This approach is focusing at polyphonic sounds and it involves four instruments, piano (P), guitar(G), saxophone(S) and violin(V). The number of instruments is not what makes this approach interesting, instead it is the fact that it can be more accurate than traditional machine learning methods, even in presence of interference using a pairwise comparison scheme and one carefully designed feature.

Setting up the experiment

During the training stage 1.000 mixtures were used. In order to set it up the signal had to be segmented and the number of instruments that was presented in each segment had to be defined. Additionally the fundamental frequency (F0) for each instrument was estimated. It should also be noted that the test set didn't use any of the instrument samples.

Feature selection and extraction

The features to be extracted depended directly on which instruments were being considered and were calculated individually for each partial. Since four instruments were being involved in the experiment, as mentioned above, there were six possible combinations. Some pairs had more similar characteristics and some were considerably different, hence, this varied the level of difficulty. A total of 54 features were considered.The feature selection for each pair aimed for the best linear separation. The following table shows the best separation accuracy for each pair of instruments:



	SV	SP	SG	PV	VG	PG1	PG2
Acc.	90%	93%	94%	90%	93%	95%	83%

Figure 3: Separation accuracy for each pair of instruments

As it is shown in figure 3, there are two features for the pair piano-guitar. Since there was no feature calculated for individual partials that could reliably separate this pair, a need for a new feature that could be more accurate came up.

Instrument identification procedure

From the setting up procedure the number of instruments for each segment and the respective fundamental frequencies are known. For each isolated partial a pairwise comparison is applied and an instrument is chosen as the winner for each pair. The same procedure is repeated for all partials related to that fundamental frequency. Then, the predominant instrument is taken as the correct one. The same is repeated for all fundamental frequencies.

Performance

The performance of this approach is exceptionally good. In the following table it is shown that when more partials are available the accuracy can reach up to 96%.

Isolated partials available	Accuracy
one	close to 91%
more than six	Up to 96%

There are three factors that play an important role for such good performance. The first one of them is the small number of instruments, so far we have seen that bigger number of instruments, results to less accuracy. Also the fact that only one database is used for all instruments makes things easier and improves performance. The last factor is that a very effective feature was found for the difficult pair of piano and acoustic guitar, this significantly improved the overall results.

Conclusions

As it is presented in the ”Performance” section for the G. Agostini, M. Longari, and E. Pollastri approach, the QDA classifier provided results that couldn't be competed by common used classifiers. The reason for this good performance seems to be that the features that were extracted form isolated tones follow a similar distribution. Statistical tests are still in process on the dataset used to validate this hypothesis. A drawback in the G. Agostini, M. Longari, and E. Pollastri approach is that the feature set used still lacks temporal descriptors of the signal. The next steps for this approach, regarding this drawback, are to introduce new features e.g. log attack slope or new timing cue schemes like the cited hmms. Additionally a new session of test using percussive sounds and sounds from live-instruments is planned.

Barbedo Tzanetakis work, proved that the pairwise comparison is an effective approach that can provide robust and accurate results. Since the same database is used for all the instruments and only four instruments are considered, the team is planning to use signals from other databases and include more instruments.

References

[1] Nicolas D. Chetry. *Computer Models for Musical Instrument Identification* (2006).

[2] G. Agostini, M. Longari, and E. Pollastri. *Content-based classication of musical instrument timbres. International Workshop on Content-Based Multimedia Indexing* 2001.

[3] Jayme Garcia Arnal Barbedo and George Tzanetakis. *Instrument identification in polyphonic music signals based on individual partials*. 2010.