Learning Instrument Identification

Lewis Guignard and Greg Kehoe

Abstract

In this project we perform instrument recognition from a mono-track audio recording. We focus on single instrument classification for eight different instruments.

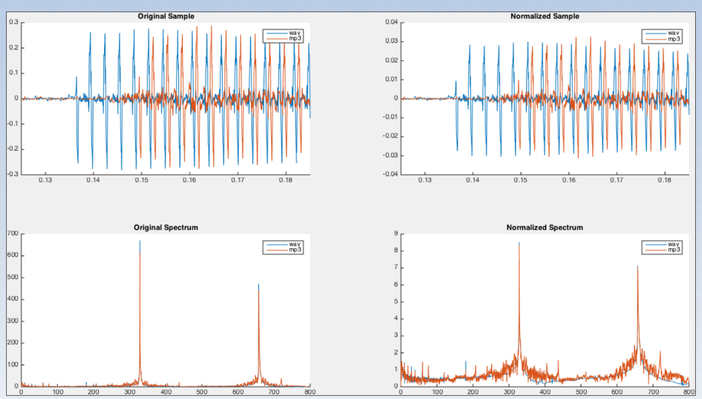
- Acoustic Guitar
- Electric Guitar
- Electric Bass Guitar
- Tenor Drum
- Bass Drum
- Snare Drum
- Cymbals
- Hi-Hat

Data Collection & Sample Quality

We used two methods of data collection, each with varying levels of quality in obtained in the sample. Our analysis shows that the spectral content is well-preserved for various bit-rates.

- Live recording
  Controlled quality for sample generation (mp3 or wav format)
- YouTube audio extraction
  Quality limited by online content
  (mp3 format at various bit rates)

$$\frac{\|wav - mp3\|}{\|wav\|} = 0.0544$$

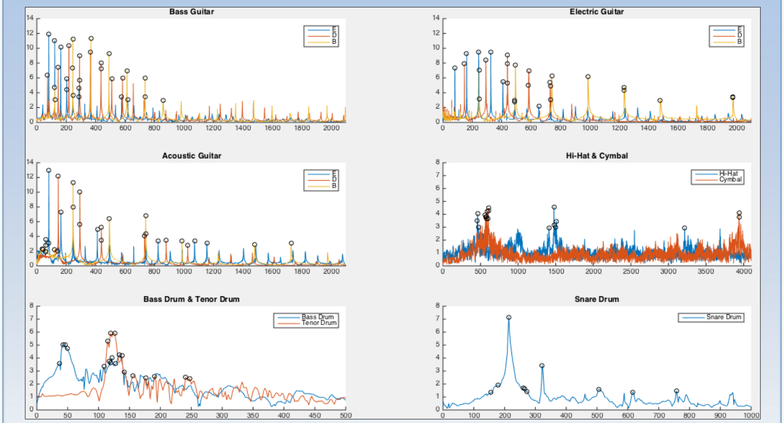


WAV vs. MP3 Format

Feature Selection

After normalizing all samples to unit energy, we extract 10 features from the DFT of each sample. Each of the n features consists of a 3-tuple, thus each example is a length-30 feature vector.

- Frequency f_n of nth largest magnitude
- Magnitude of peak at frequency f_n
- Power contribution of frequency f_n



Feature Selection and Instrument Comparison

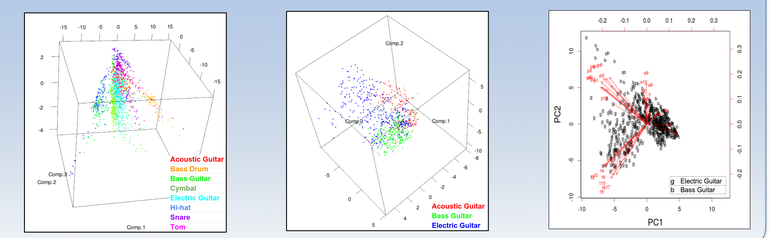
K-Means

Using k-means, we attempted to group each instrument into its own cluster. We discovered that certain instruments were not discernable by using k-means, while other instruments are easily separable. Intuitively, instruments that are significantly different in their acoustic qualities are more easily separated into distinct clusters.

- Electric Guitar and Cymbals
  93.2% correct labeling
- Electric Guitar and Tenor Drum
  58.4% correct labeling
- Electric, Bass, and Acoustic Guitar
  No discernable mapping from clusters to instrument class

Principal Component Analysis

Using the first three principal components, the data set appears to be separable, although the three guitars are very similar and more difficult to classify given our feature selection.



SVM Classification

We ran SVM for all instruments (8 classes) and for the Electric Guitar, Bass Guitar, and Acoustic guitars only (3 classes). For multi-class situations we used a set of one-versus-one classifiers.

For each method, we used three kernels, trained on 80% of a random permutation the data, and evaluated training and test error. We iterated over several settings of hyper-parameter values to determine the lowest training error then used the optimal hyper-parameters to run 10-fold cross-validation to estimate the test error of each model.

All Instruments (1440)

Kernel	**Linear**	**Radial (Gaussian)**	**Polynomial**
Cost parameter	1	10	0.1
Gamma	n/a	0.5	1
Degree	n/a	n/a	2
Training Error	8.00%	8.00%	8.00%
Test Error	20.5	30	17.9
Support Vectors	713	1105	681
10-fold cv error %	**21.5**	**30.9**	**17.8**

Guitars (771)

Kernel	**Linear**	**Radial (Gaussian)**	**Polynomial**
Cost Parameter	5	10	0.1
Gamma	n/a	0.5	0.5
Degree	n/a	n/a	3
Training Error	6.5	6.5	6.5
Test Error	17.4	3.6	6
Support Vectors	188	604	299
10-fold cv error %	**14.4**	**38.2**	**13.1**