Classification and visualization of sounds
Session 2: Musical sounds

Mathieu Lagrange

December 8, 2015

This document, the Matlab scripts and the datasets are available here:
http://www.irccyn.ec-nantes.fr/~lagrange/teaching/pds/tpVisualization

By convention, the Matlab functions that are available as built-in or in the rastamat directory are in bold font.

The "musicGenre" dataset is composed of 10 pairs of music clips. Each pair belong to a given genre: Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae, Rock.

1 Napping

1. Download and decompress the "musicGenre" dataset
2. Use the **napping** function to explore this dataset (clicking nearby a dot allows you to listen to the sound)
3. organize the sounds on the 2D plane in order to have similar sounds that are close to each other and dissimilar sounds that distant.
4. the locations that you set are stored in Comma Separated Value (CSV) file. Copy it to a file with a name containing the name of the dataset and your name.

2 Description

We aim at describing each sound with a set of descriptors, called features. Each features shall be computed on successive overlapping blocks.

1. decide which block size and overlapping factor to use
2. implement the spectral centroid feature using the **spectrogram** function
3. implement the Mel Frequency Cepstral Coefficients (MFCC)s. To this end, use the **audspec** and **spec2cep** functions. For each function, display and describe the second output argument.
4. implement the beat histogram feature in order to model the rhythmic aspects of music. The beat histogram shall be computed as follows: split a spectrogram into 3 frequency bands: low, medium, high. For each band, perform a Fourier transform in order to identify the 3 dominant periodicities and store them in an histogram.

3 Visualization

1. Perform a Principal Component Analysis (PCA) of the averaged MFCC features
2. Get the coordinate of the sounds into a 2D place that maximizes the displayed feature variance
3. Use the **napping** function to display this projection.
4. Compare qualitatively and quantitatively the organization of the sounds in the "acoustical" plane and in the "human" one using the **daniels** function.

4 Extra: classification

We aim at classifying the sounds into their corresponding classes. To do so, we consider a 1 Nearest Neighbor (1-NN) approach using the Euclidean distance. Perform the pooling operation over several texture windows. Prediction is done by majority voting over the several texture window. An example, suppose that your sound is of 12 frames and your texture window is of size 4, the 12 frames are averaged 4 by 4 to give 3 features. Those 3 features a compared to the computed features of the other sounds of the database. Suppose the predicted classes are Rock, Pop, Rock, a majority vote would choose for the Rock class.

1. For each feature, compute the prediction accuracy, that is the number of sound for which the closest sound is of the same class.
2. Does a feature normalization improves the results ?
3. For the MFCCs, shall the 1 coefficient be kept ?
4. Which features gives the best accuracy ?
5. Is it beneficial to combine features ?

5 Report

Please write a report using your favorite word processing tool and output a pdf file. The report shall have for each question a brief description about the way things have been done and some discussion about the resulting behavior.

Send an archive containing the report, the code and the csv files no later than an hour after the end of the session to `mathieu.lagrange@cnrs.fr`, with the `[PDS]` flag within the title of the message.

A Useful commands

A.1 Miscellaneous

- hist : histogram
- repmat : matrix replication
- imagesc : scaled matrix display

A.2 Distance

- pdist : pair wise distance computation
- squareform : convert output of pdist from vector to matrix

A.3 Documentation

- doc "command" : documentation of "command"
- lookfor "keyword" : show commands with "keyword" in the description