Classification and visualization of sounds
Session 1: Environmental sounds

Mathieu Lagrange

December 9, 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 "environmentalSound" dataset is made of 8 pairs of sounds. Each pair belong to a given category: Bird, Dog, Door, Glass, Harp, Keyboard, Ping Pong, Traffic.

1 Napping

1. Download and decompress the "environmentalSound" 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 feature shall be computed on successive overlapping blocks and averaged.

1. decide which block size and overlapping factor to use
2. implement the Zero Crossing Rate (ZCR) feature
3. implement the spectral flatness feature. Use the **spectrogram** function.
4. implement the tonal power ratio. to this end the tonal part of the frame is defined as local maxima of the magnitude spectrum that are above a given threshold

3 Visualization

1. Get the coordinate of the sounds into a 2D space using 2 of the features implemented. Generate a 'csv' file with per line, the 2 features and the color as 3 RGB values.
2. Use the **napping** function to display this projection.

3. Compare qualitatively the organization of the sounds in the "acoustical" plane and in the "human" one. Perform a quantitative analysis using a normalized correlation of vectorized version of the similarity matrices:

$$c(a,b) = \frac{ab'}{\sqrt{aa'.bb'}} \tag{1}$$

4. Find the combination of features that correlates best with your reference.

5. (Extra) implement the Mantel test in order to compute an alternative correlation indicator.

4 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. For each sound, the feature is the average in time of the feature set. That is, if a sound is analyzed into 23 frames, each represented using 2 coefficients, the resulting feature is a vector of size 2.

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. Which features gives the best accuracy ?

4. Is it beneficial to combine features ?

5 Report

The report will be due at the end of the second session.

Please write the 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.

Please add to the report the csv file that you created in Section 1.

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