Report 1 – Feature extraction

Objective

During this lab, we will understand the different steps for extracting the features of the sound and how to differentiate the sounds.

Enframe

The purpose of this step is to cut the original sample in many smaller ones.

Here the shift is smaller (10ms) than the length of the window (20ms), part of the original will be in more than one window.

Calculation of the length and the shift in samples :

Sample rate (Sp) : 20 kHz

Time between two samples (Tp) : $\frac{1}{Sp} = 5\text{e-}5$

Length of the window : $0.02/Tp = 400 \ \ samples$

Length of the shift : $0.01/Tp = 200 \ \ samples$

Pre-emphasis

This function compensate the 6db/octave that are dropped due to the radiation at the lips.

The function is like $y[n]=x[n]-\alpha x[n-1]$

The coefficents are A = 1 and B = $[1 \ \ -\alpha]$

Hamming Window

The hammig window will reduce the extremities of the window and keep emphasize the center of the window. This is really important for improving the values of the FFT. This will reduce the high level harmonics and focus on the main frequencies.

Fast Fourier Transform

According to the sampling theorem :

$$f_{max}=\frac{f_s}{2}=\frac{20000}{2}=10\,kHz$$

Mel filterbank log spectrum

The filters is composed of 40 filters. The first triangle is only keeping low frequencies. The next filters are slowly shifting to the higher frequencies. However, the amplitude of the triangle is decreasing.

Result

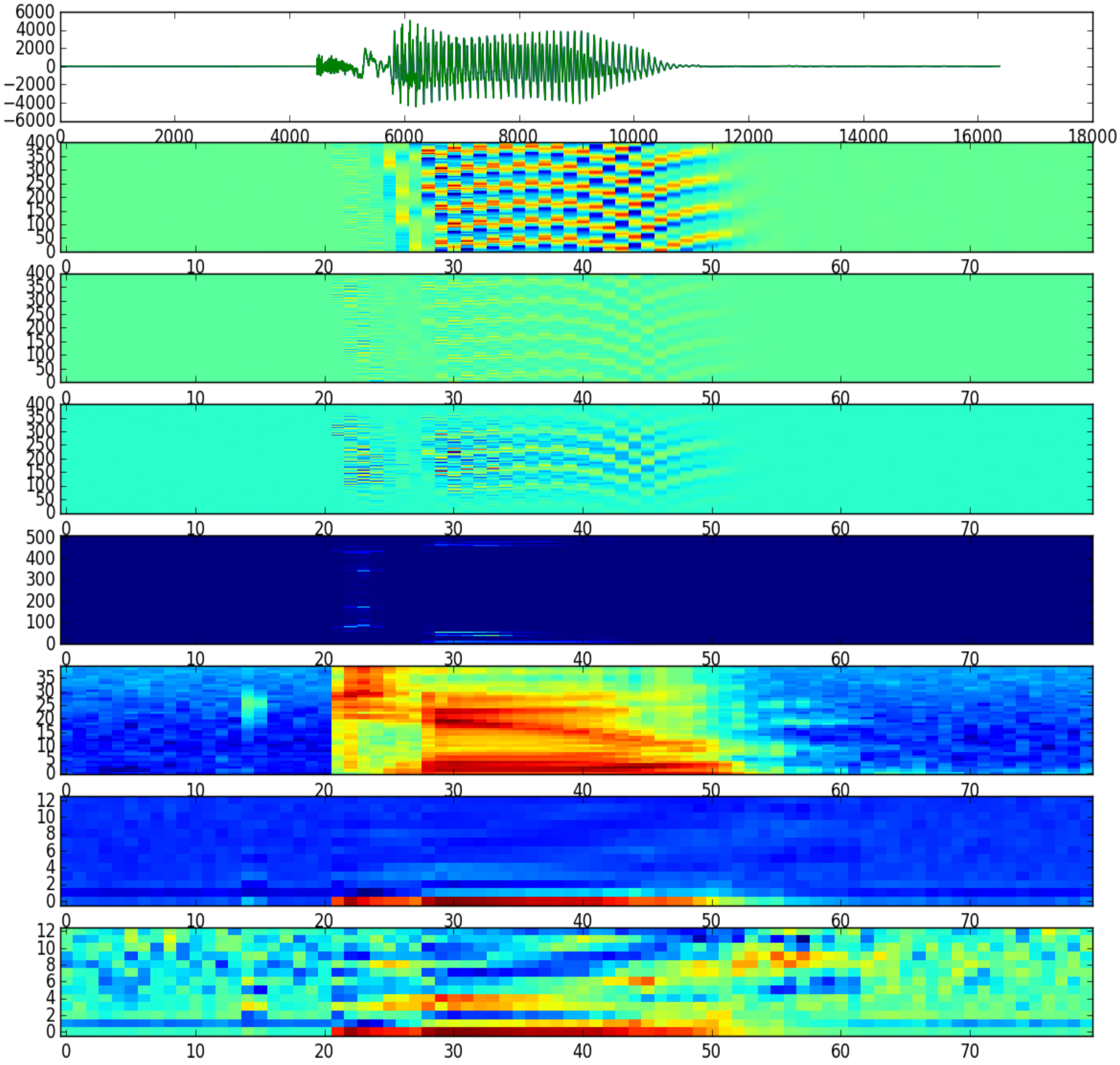
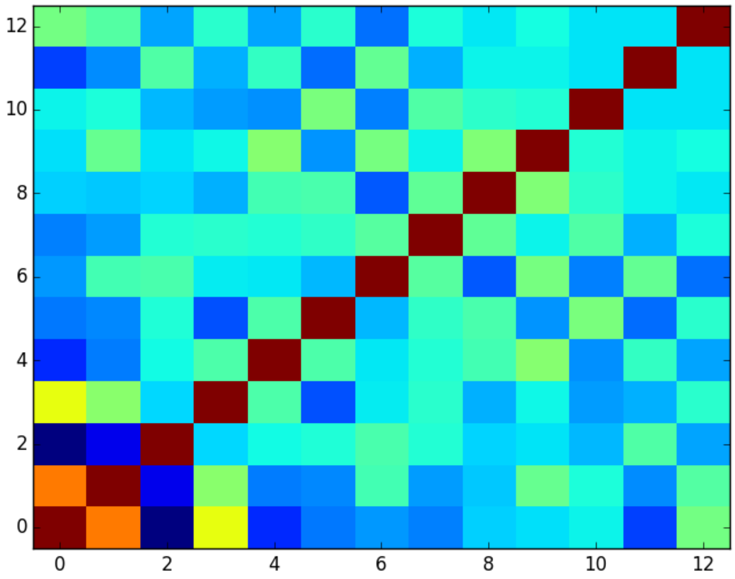


Illustration 1: Differents steps for extracting the features of the example

Feature Correlation



We can see that the correlation is really strong on the diagonal and really weak elsewhere. Those features are really good to use because each vector is giving a lot of new information.

Illustration 2: Correlation between Mfcc features

Using Mspec features is a really bad idea since the correlation is really strong between vectors that are neighboor.

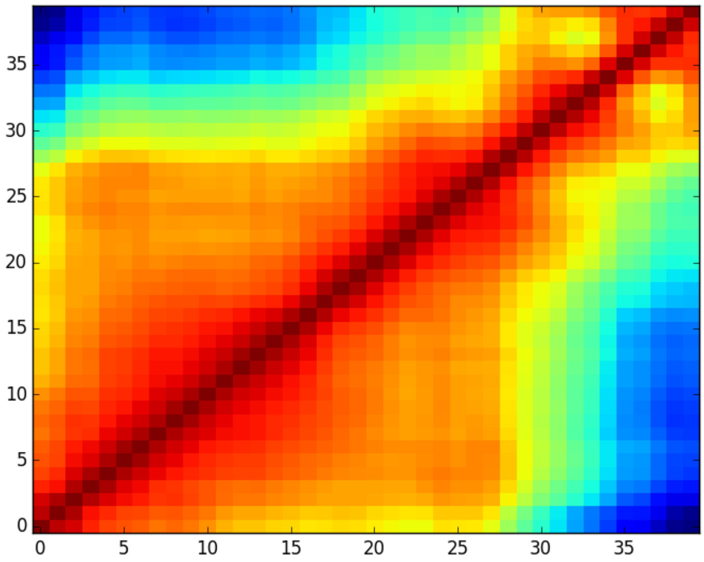


Illustration 3: Correlation between Mspec features

Comparing Utterances

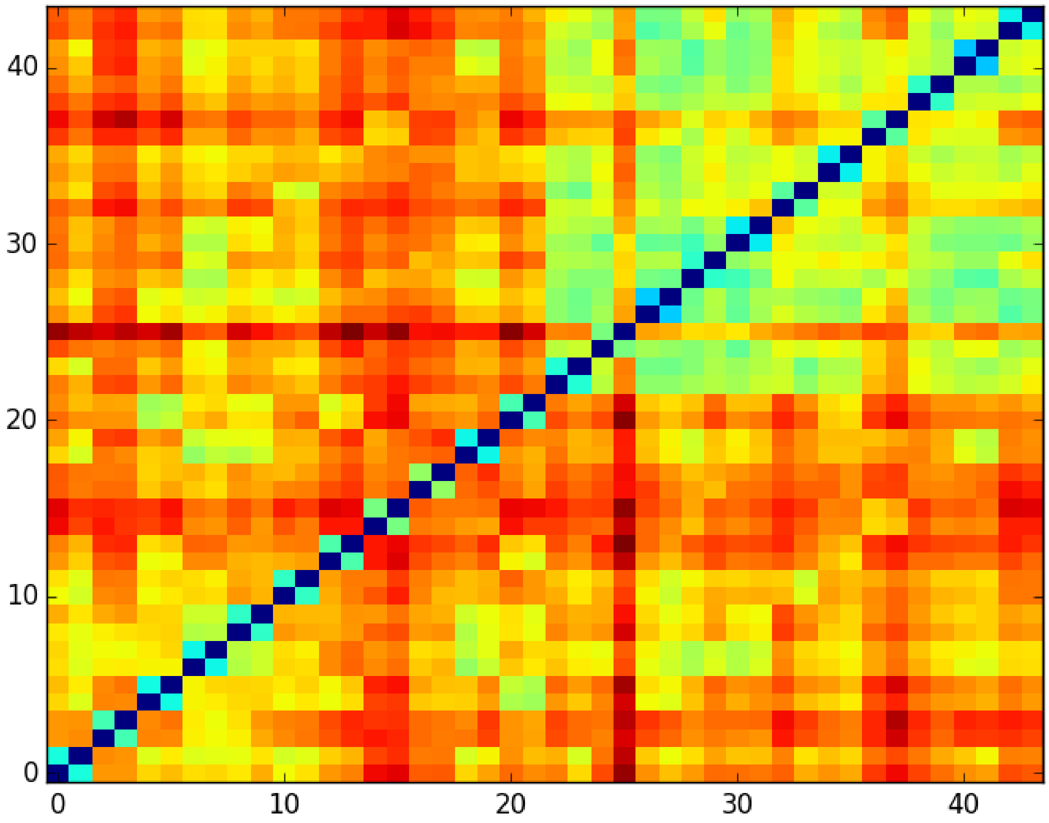
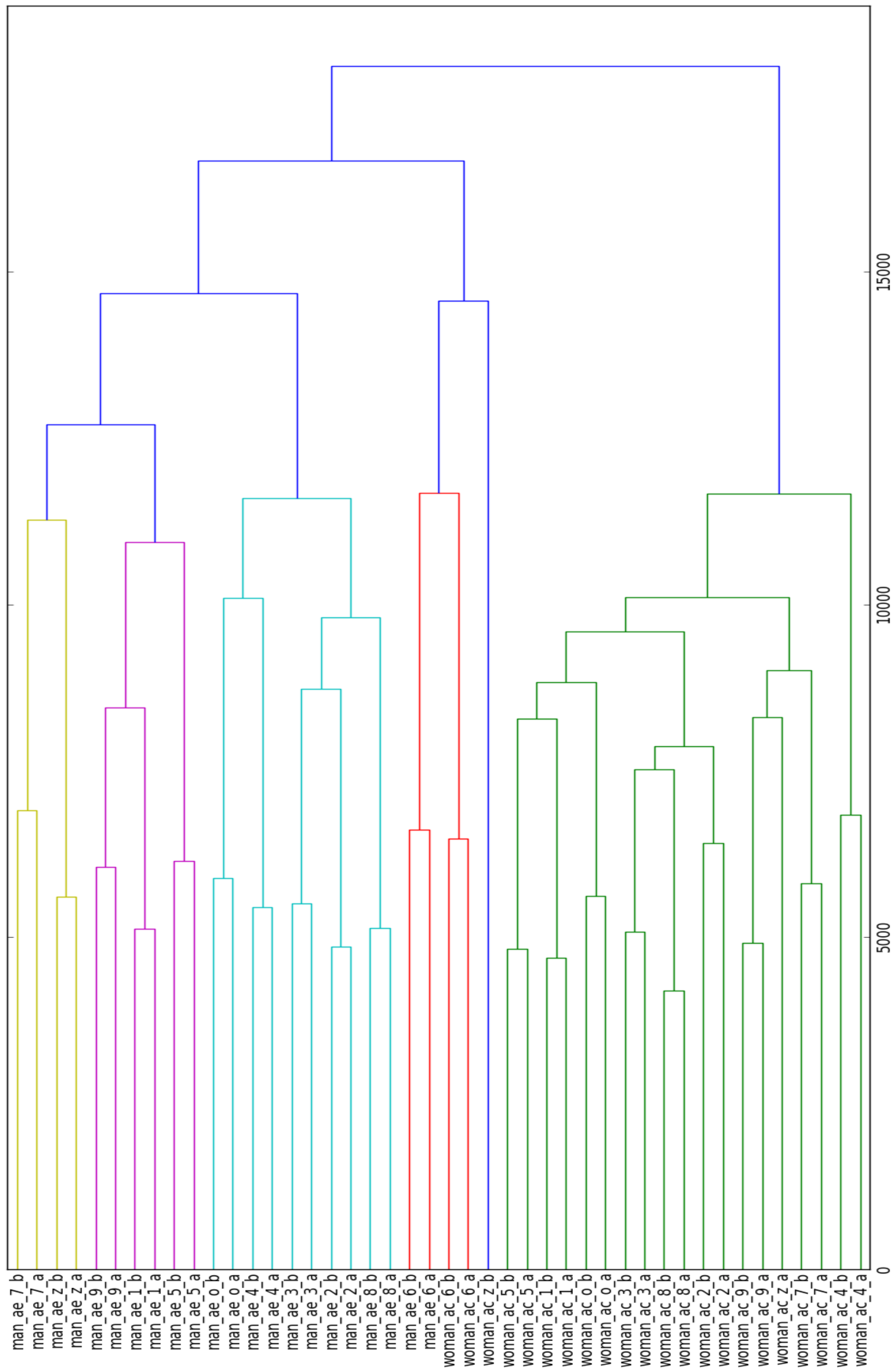


Illustration 4: Distance between samplings of tidigits

We can see that the distance is really small between a « digit » said from a girl and a man. Thus it will be harder to recognize them.

For the samples 0 to 22 to every other samples the distance is more emphasized so it will be easier to recognize them. The distance between the samples 22-44, that are mostly numbers, have a smaller distance between them. The system has more difficulties to differenciate numbers.

Dendrogram


0
5000
10000
15000
man_ae_7_b
man_ae_7_a
man_ae_z_b
man_ae_z_a
man_ae_9_b
man_ae_9_a
man_ae_1_b
man_ae_1_a
man_ae_5_b
man_ae_5_a
man_ae_o_b
man_ae_o_a
man_ae_4_b
man_ae_4_a
man_ae_3_b
man_ae_3_a
man_ae_2_b
man_ae_2_a
man_ae_8_b
man_ae_8_a
man_ae_6_b
man_ae_6_a
woman_ac_6_b
woman_ac_6_a
woman_ac_z_b
woman_ac_5_b
woman_ac_5_a
woman_ac_1_b
woman_ac_1_a
woman_ac_o_b
woman_ac_o_a
woman_ac_3_b
woman_ac_3_a
woman_ac_8_b
woman_ac_8_a
woman_ac_2_b
woman_ac_2_a
woman_ac_9_b
woman_ac_9_a
woman_ac_z_a
woman_ac_7_b
woman_ac_7_a
woman_ac_4_b
woman_ac_4_a