UNIVERSITY OF CRETE
DEPARTMENT OF COMPUTER SCIENCE
Multimedia Technology
Spring 2018

7th assignment

The objective is to segment audio signals and then to discriminate the resulting segments in speech and music. The only feature to be used for achieving these tasks is the signal amplitude. Specifically, the average amplitude will be computed in non-overlapping frames of 20 msec.

Initially, the power values of signal x will be smoothed using a linear system implemented using Matlab function

```
filtfilt(b, [1 -b], x.^2)
```

with $b = 0.05$. Then the instantaneous amplitude is estimated by taking the square root of the smoothed signal power and the average amplitude in each frame is calculated. It is requested to plot the average amplitude versus time in sec. The average amplitude is also smoothed with the above filter, and the values $\{A(n), n = 1, \ldots, N_B\}$ are obtained, where N_B is the number of frames. The unsigned differences $D(n) = |A(n+1) - A(n-1)|$ are then calculated. Plot $D(n)$ versus time in sec. Identify important differences in average amplitude by finding the maxima of the difference $D(n)$ that are more than twice the mean difference and are at least distanced 0.5 sec. In this way the audio signal is segmented. The maxima instants, where the difference is important, are possibly changes from voice to music or vice versa.

In order to classify the segments resulting from the detection of the difference maxima, the mean value and the standard deviation for the average amplitude will be measured for each segment. The decision is based on the ratio of the mean value to the standard deviation. If the ratio is less than 2.25, it is decided that it is a voice signal, otherwise it is a music signal. Discriminate the whole music and voice segments, and obtain their description with the ratio of the mean value to the standard deviation of the average amplitude and with the same ratio for the average zero-crossing rate calculated on the same frames. The average zero-crossing ratio is given by

$$z = \frac{1}{2(N-1)} \sum_{n=1}^{N-1} |\text{sign}(x(n+1)) - \text{sign}(x(n))|.$$

Results should be given for signal *TestSegment.wav*. The work report, the results and the conclusions should be given in HTML.

Useful Matlab functions : audioread, filtfilt, findpeaks, mean, std.