

SGN-14007 Introduction to Audio Processing

Exercise 5

Week 48

In this exercise, you will implement one of the commonly used features in audio signal processing - **MFCC**. There are two questions with 1 point for each summing up to 2 points. Bonus question is optional worthing 0.5 points. The submission should consist of a **report** of your observations and the python **code**. Alternatively, both can be combined in a Jupyter notebook as well.

Question 1 : Mel filterbank (1 point)

(a) Load the given audio file (audio.wav) to get the audio signal and its sampling rate. Hint: *librosa.load('audio.wav')*

(b) Create a mel filterbank with the following settings:

- Sampling rate obtained from (a)
- $N_{\text{fft}} = 512$
- $N_{\text{mel}} = 40$

Hint: *librosa.filters.mel(sr, n_fft, n_mel)*

(c) Plot the mel filterbank and report your observations. Hint: *librosa.display.specshow*

Question 2 : MFCC (1 point)

(a) Pre-emphasis the audio signal (audio.wav) with the given Equation $y(t) = x(t) - \alpha x(t-1)$, $\alpha = 0.97$.

Hint: *np.append(x[0], x[1:] - α * x[:-1])*

(b) In the stft loop you implemented earlier, for each frame, do the following:

- Window each frame (using hamming window) Hint: *signal.hamming*
- Calculate the fft
- Collect the power spectrum (you will get power spectrum)
- Multiply it with mel filterbank you created in Question 1 (you will get mel spectrum)
- Take log operation (you will get logarithmic mel spectrum) Hint: *20 * np.log10*
- DCT (Finally, you will get MFCC)

(c) Plot logarithmic power spectrums, mel spectrums, logarithmic mel spectrums, MFCC.

(d) Implement MFCC using librosa and compare with yours. Report your observations.

Bonus Question : Implement your own mel filterbank with the same setting as Question 1 (0.5 point)

In Question 1, you used librosa library to get mel filterbank. Now try to implement by yourself!

Hint:

- convert hz to mel using Equation $F_{mel} = 2595 \log_{10}(1 + f/700)$
- uniformly distributed on the mel scale
- convert mel to hz back using the inverse of Equation above
- Mel filter of band m:
 1. starts at 0 amplitude at $F_{mel,c}(m-1)$
 2. has maximum amplitude 1 at $F_{mel,c}(m)$
 3. decays to zero at $F_{mel,c}(m+1)$

You can follow the equation below.

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k < f(m) \\ 1 & k = f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases}$$

- to have a flat spectrum in mel domain for a DFT magnitude spectrum, the mel bands need to be scaled

Plot it and compare with librosa implementation. Report your observations.