Tutorial 5 - CS4347- 17 Feb 2017

Today's tutorial is about extracting perceptual features using python. Perceptual features are features which attempt to capture the human *perception* of sound. For example, the cochlea (located in our ears) can not discern the difference between two very close frequencies. In our previous assignment on spectral features, we extracted these frequencies without considering how (and if) they are detectable by humans. This tutorial will be on one type of perceptual feature: **mel-frequency cepstral coefficients (**MFCC). But first, take a seat, *we need to talk*:

In sound processing, the **mel-frequency cepstrum** (**MFC**) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a non-linear mel scale of frequency. Got it? No? Ok lets keep going… :)

Filters
A filter is a device or process that removes some unwanted components or features from a signal spectrum. You can think of it like a "window" in the frequency domain.

The Mel scale
The Mel scale relates *perceived frequency*, to its actual measured frequency. The mel-scale is a non-linear scale that is adapted to the non-linear audio perception of the human auditory system. Humans are much better at discerning small changes in low frequencies than they are at high frequencies. This scale allows us to match more closely to what humans hear.

Mel-spaced filter banks
The cochlea can not discern the difference between two close frequencies and this effect becomes even more pronounced at higher frequencies. To give us an an idea of how much *energy* exists in various frequency regions, we take groups of our FFT bins and sum them. This is performed using Mel-spaced filter banks, a set of 20-40 (26 is standard) triangular filters that we apply to the FFT bin energies.

The first filter is very narrow and gives an indication of how much energy exists near 0 Hz. As the frequencies get higher, our filters get wider as we become less concerned about energy variations at higher frequencies (see Figure 1). Each filter of the filter bank is non-zero for a certain section of the spectrum. To calculate filter bank energies we multiply each filter with the power spectrum, then add up the coefficients. Once this is performed we are left with 26 numbers (for 26 filters) that give us an indication of how much energy is in each of the filters.

Once we have the filter bank energies, we take the **logarithm** of them. This is also motivated by human hearing: we don't hear loudness on a linear scale. Generally, to double the *perceived volume* of a sound we need to put eight times as much energy into it. This means that large variations in energy may not sound all that different if the sound is loud to begin with.

The final step is to compute the **discrete cosine transform (DCT)** of the log filter bank energies. This is done because our filters are all overlapping, and thus the filter bank energies are correlated with each other. The DCT performs two tasks for us here:

1. It decorrelates these energies.
2. It can reduce the number of dimensions of our filter bank because it tends to compact most of the energy of the spectrum in the first few coefficients.

Neat, you have calculated the MFCC, try plotting them!

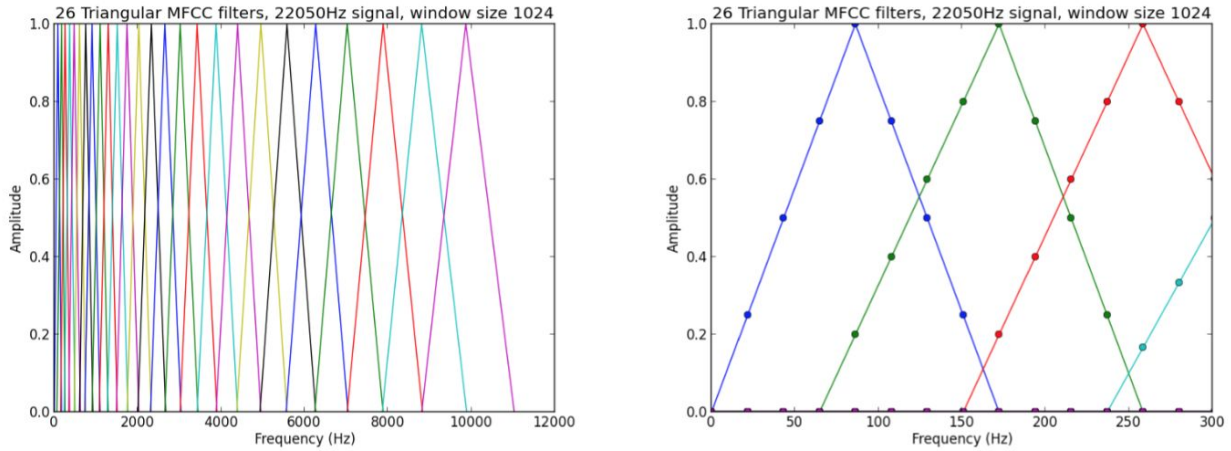

Figure 1: The overall range of the triangular windows, and the triangular windows from 0 to 300 Hz

Great summary here:
http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

Your missions this week:

1. Read through and try to understand about MFCCs
2. Start the assignment!
3. Bonus: Plot your resulting MFC coefficients. How do they compare to your FFT from assignment 2?

Some questions to think on:

1. Why do we use the Mel-scale?
2. How is DCT different to FFT?
3. What are some other applications of DCT?
4. What other perceptual features did you learn about in the lecture and how can they be extracted from audio?

As always, your journey doesn't stop here. Have a look at these other interesting links:

https://en.wikipedia.org/wiki/Mel-frequency_cepstrum

https://en.wikipedia.org/wiki/Discrete_cosine_transform#Informal_overview

http://dsp.stackexchange.com/questions/8866/what-is-the-purpose-of-the-log-when-computing-the-mfcc

http://dsp.stackexchange.com/questions/15938/is-this-a-correct-interpretation-of-the-dct-step-in-mfcc-calculation

Ask your friendly neighbourhood TAs if you are having any problems, and remember google/baidu/bing? is your friend!

Next time on CS4347: We are going to look at extracting… wait... you have a mid-term test. Go study!