Assignment 2

Yang Gao
yanggao@kth.se

May 2018

1 Plot of Sound Signal Over Time Domain

The code used to generate following plot can be found in `plottime.m`]
First, the music signal plot over time domain

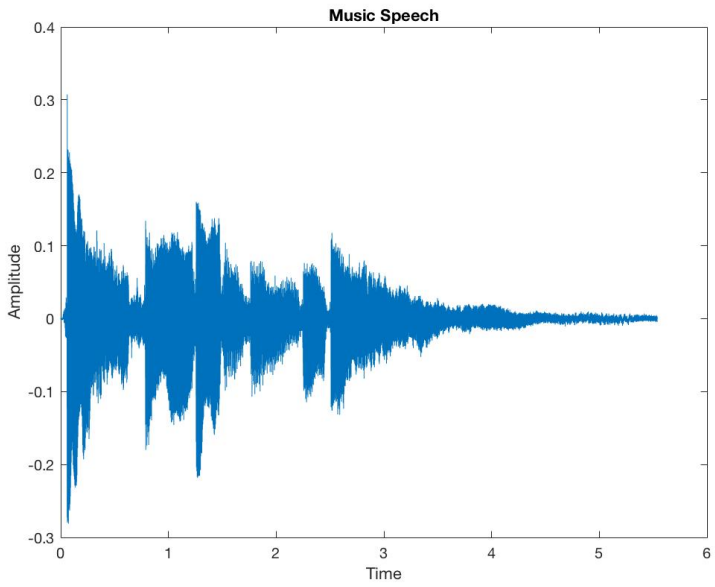

Figure 1: Music signal plot over time

Second, the female signal plot over the time domain

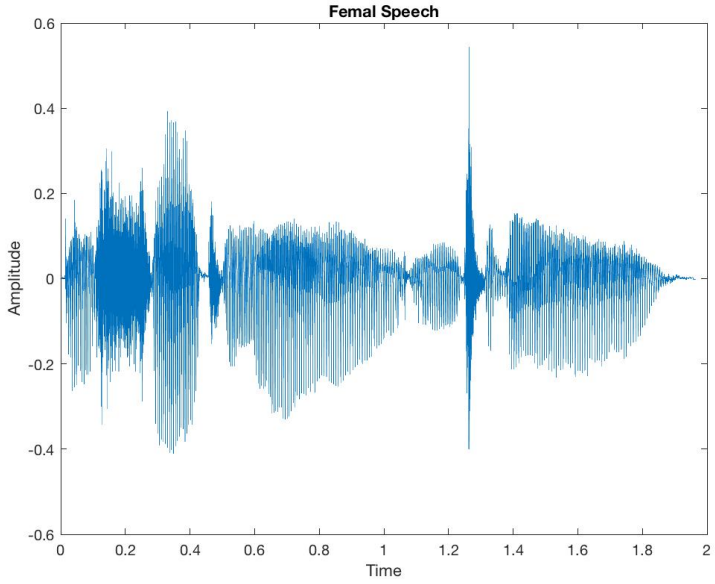

Figure 2: Female signal plot over time

The voiced sections of the signal can be locate on the plot where the oscillatory with some sort of pattern

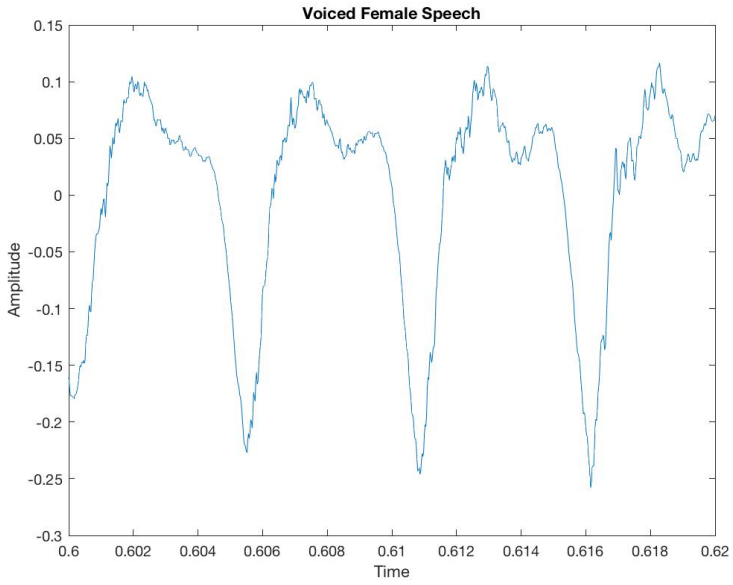

Figure 3: Voiced section of the female speech over time

The unvoiced section of the signal can be locate on the plot where the oscillatory without any sort of pattern

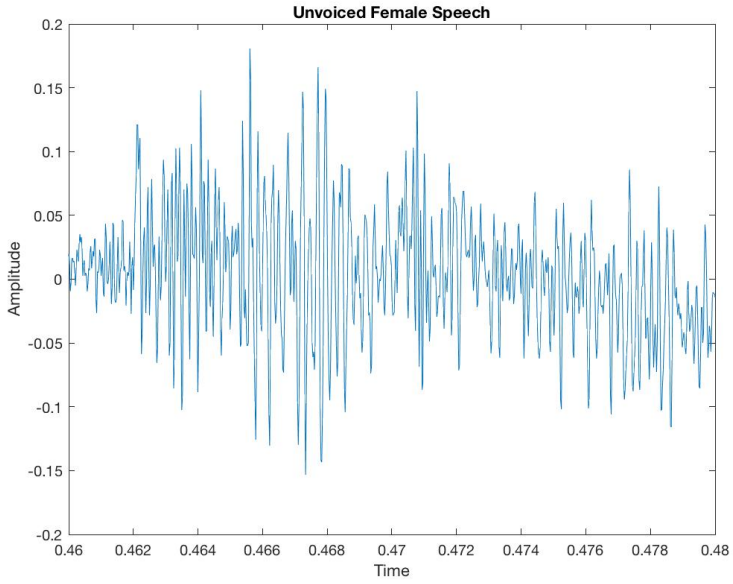

Figure 4: Unvoiced section of the female speech over time

2 Spectrogram

For the music signal, the harmonic are highlighted on the plot the code generated the plot can be found at `spectplot.m`]

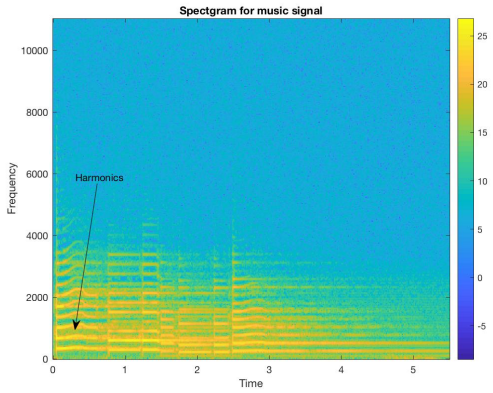

Figure 5: Spectrogram for the music signal

For the female speech, it is possible to distinguish between voiced and unvoiced section from the plot, unvoiced section have no concentration on certain frequent;However, voiced section have an concentration on a certain frequency et. The voiced and unvoiced segment are highlighted on the diagram below

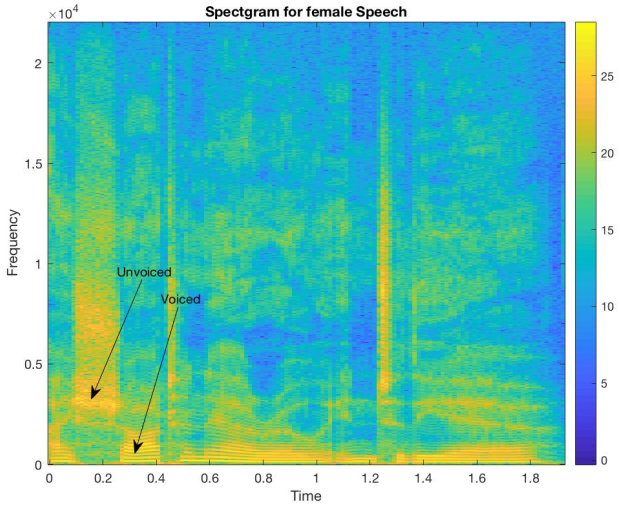

Figure 6: Spectrogram for the female signal

3 Cepstrogram

The plots below show the Spectrogram and Cepstrogram(normalized) for the music signal. the code for this section can be found at `ceptplot.m`

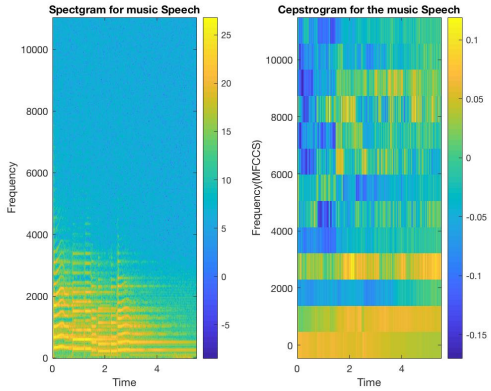

Figure 7: Spectrogram for the female signal

Then, the cepstrogram of male and female speech samples are plotted together in order to compare them in the same phase

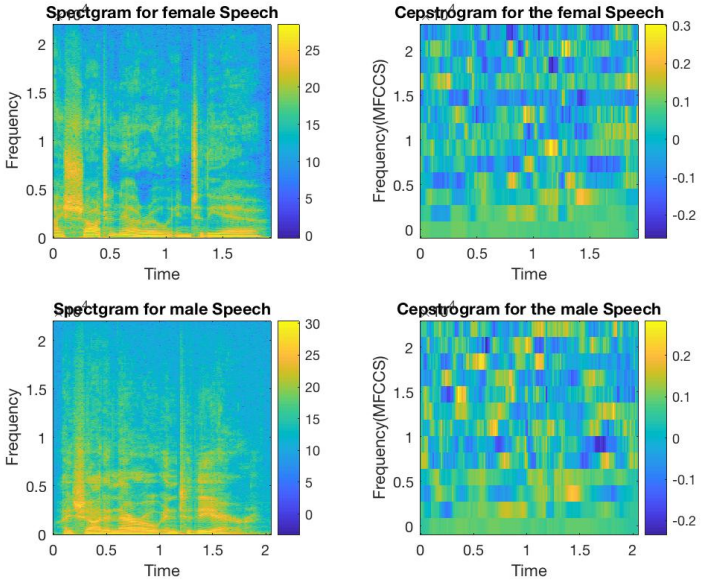

Figure 8: Spectrogram for the female signal

4 Correlation matrices

The following plot used the code `corr.m` to plot the correlation for the female speech sample

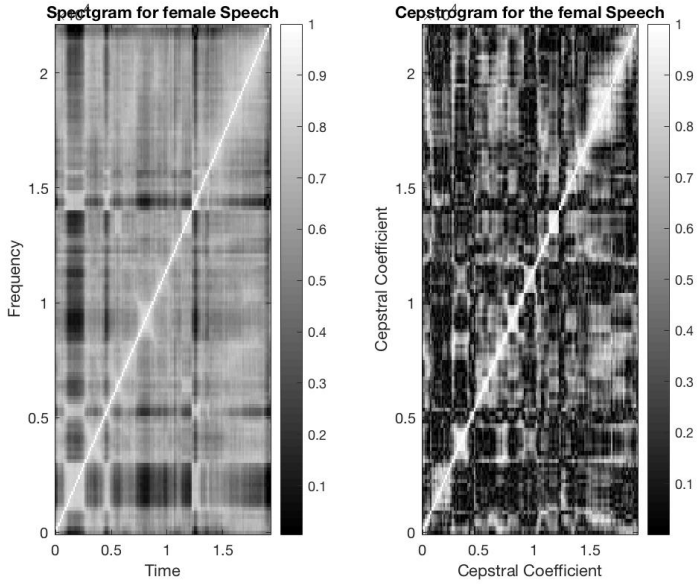

Figure 9: Spectrogram for the female signal

5 Questions

a) Which representation do you think is the easiest for you, as a human, to interpret, and why?

In my opinion,the Spectrogram is much easier for me to interpret. Since it has a very clear reflection on the signal. the Spectrogram give a trend of the signal. It also very easy to distinguish between voiced and unvoiced sections

b) Can you see that they represent the same phrase? Could a computer discover this? Why/why not?

Even though male and female have different ascents and frequent range. It is still possible for human to recognize that they are represent the same phrase since there are many similarity between them such as the trends and the concentration on certain frequency.
However, I don't think a computer can make the decision that those two Spectrogram represent the same phrase since the different in one region is too small for a computer to recognize them as a same phrase

c) Can you see that they represent the same phrase now? What about a computer?

It's almost impossible for me to distinguish those to cepstrogram since it's harder for me to find a similar pattern between them. However, I think it would be easier for computer to recognize that they are represent the same phrase since the cepstrogram divided near information into chunks which make it easier for computer to make a decision

d) Which matrix, spectral or cepstral, looks the most diagonal to you?

The cepstrogram looks more diagonal to me since its pattern are more clear to follow

6 Some thoughts on the possibility of confusing the MFCC representation

Can you think of a case where two utterances have noticeable differences to a human listener, and may come with different interpretations or connotations but still have very similar MFCCs

The MFCC first translate the wave into spectral envelope field. which make us lose lots of information about the pitch. in such case, MFCC has trouble when analysis a same word with different pitch. for example, "I am scared" has a different with "I am scared?"

What about the opposite situation: are there two signals that sound very similar to human, but have substantially different MFccs;

cepstrogram analysis have divide the sound signal into low and high frequency domain, so that when a sound contain both low and high frequency signal(Which sound very noise to human). MFCCs can easier distinguish between them.