Feature Extraction

EQ2341 Pattern Recognition and Machine Learning, Assignment 2

Jingxuan Mao jmao@kth.se
Yuqi Zheng yuqizh@kth.se

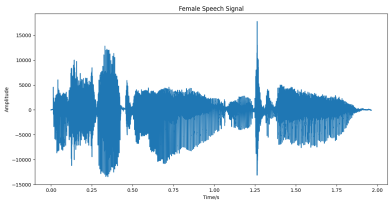
April 29, 2022

1 Introduction

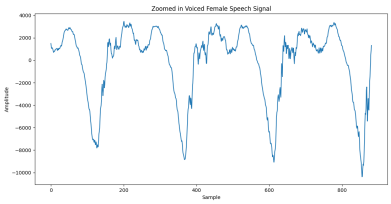
In this assignment, we look into two techniques to extract features of the sound signals: discrete Fourier transform (DFT) and mel frequency cepstrum coefficients (MFCC). We plot and compare the spectrograms and cepstrograms of the female speech signal, male speech signal and the music signal. Also, we analyze them based on the speech and acoustic principles.

2 Sound Signals

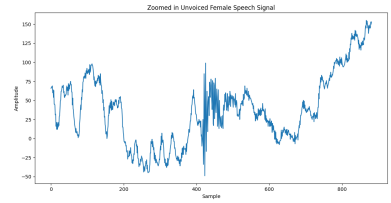
In this section, we look into the female speech signal and the music signal. We plot the two signals and zoom in on a range of 20 ms, i.e., $0.02f_s$ samples. As shown in Figure 1, we can see that for the voiced female speech signal segment, there is a clear periodic oscillation, which corresponds to a certain frequency that humans perceive while it is noisy for the unvoiced segment. Figure 2 shows the music signal where we could also observe the wavy appearances.



(a) Female Speech Signal

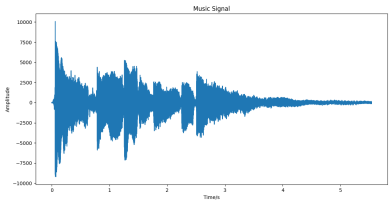


(b) Voiced Female Speech Segment

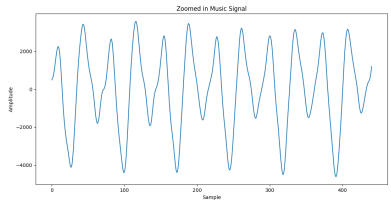


(c) Unvoiced Female Speech Segment

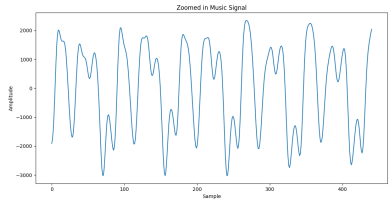
Figure 1: Female Speech Signal



(a) Music Signal



(b) Music Signal Segment



(c) Music Signal Segment

Figure 2: Music Signal

3 The Fourier Transform

3.1 Spectrograms of female and music signal

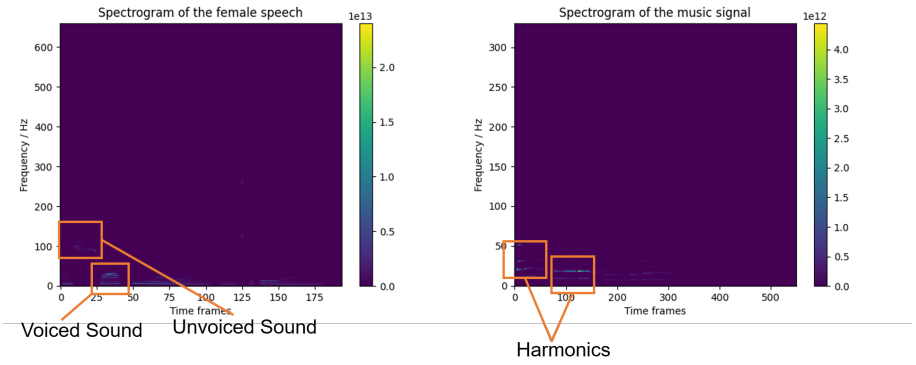


Figure 3: The Spectrograms of female and music signal

In this section, we calculate and plot the spectrograms of female and music signal. For both signals, we implemented Hanning window and set the frame size and the step size of FFT to the length of 30ms and 10ms, respectively. Also, we denoted the voiced and unvoiced segments in the speech as well as the harmonics in the music signal as shown in Figure 3.

3.2 Logged spectrograms of female and music signal

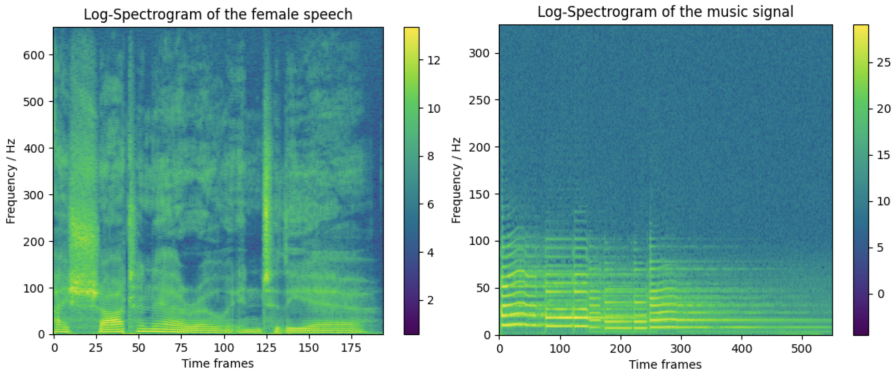


Figure 4: The Logged spectrograms of female and music signal

We calculated the logged spectral coefficients of female speech and music signal and plot them in Figure 4. It can be seen that the change of colors is more distinguishable in the logged version of spectrograms.
In the spectrogram, the horizontal stripes are the harmonics in the signal. Those spread over almost all frequencies correspond to the unvoiced sounds in the speech. We can also see that the intensity of low frequency harmonics is higher.

4 MFCCs

4.1 Comparison between spectrogram and cepstrogram

We plot the logged spectrograms and the normalized cepstrograms (with zero mean and unit variance) of female and music signal. For the cepstrogram of female speech signal, we implemented Hanning window and set the length of the analytical window and the step as 30ms and 10ms respectively, with 30 filters in filterbank and FFT size of 1323. For that of music speech, the parameters are same except the FFT size is changed to 662. The results are shown in Figure 5 and Figure 6.

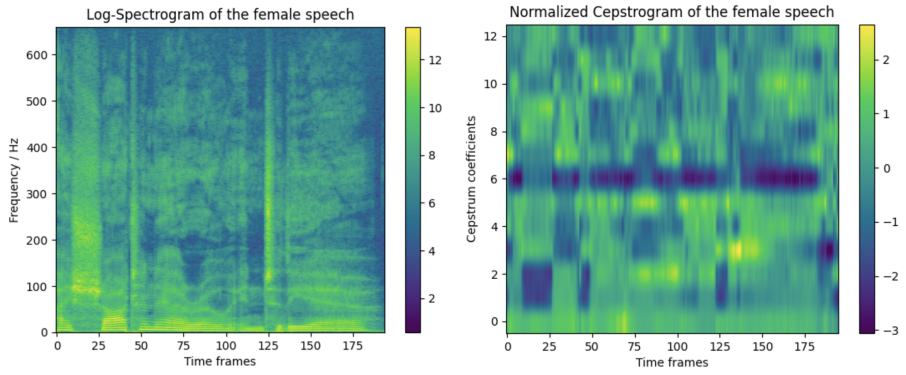


Figure 5: The Logged spectrogram and normalized cepstrogram of female speech

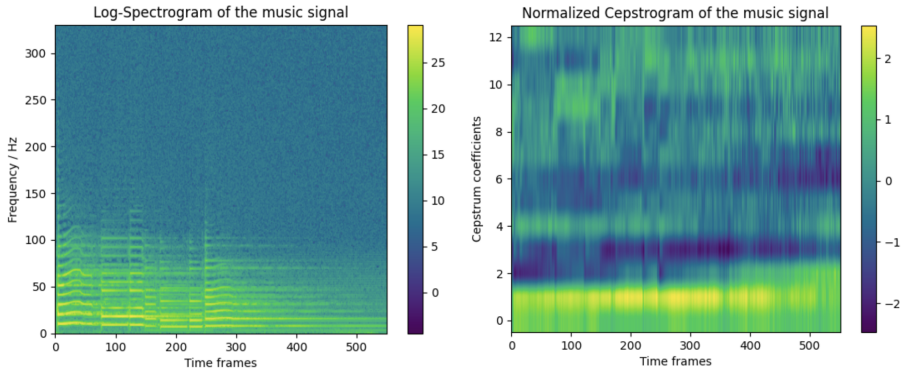


Figure 6: The Logged spectrogram and normalized cepstrogram of music signal

It can be seen that the voiced and unvoiced region of the speech is more recognizable in the spectrogram, while in cepstrogram it is easier to distinguish between different cepstral bands.

4.2 Comparison between female and male speech

In this section, we plot the logged spectrogram and the normalized cepstrogram of female and male speech signal. The content of these two audio are exactly the same, so all the parameters of the spectrogram and cepstrogram are also set to same values as that of female speech mentioned in Section 4.1. The results are shown in Figure 7 and Figure 8.

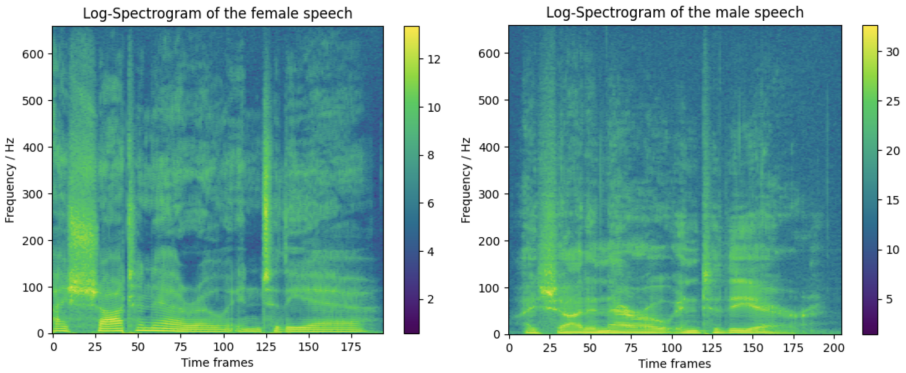


Figure 7: The Logged spectrogram of female and male speech

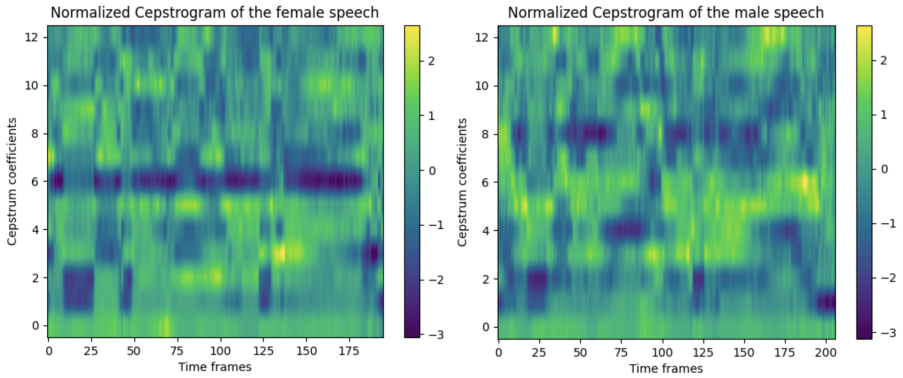


Figure 8: The normalized cepstrogram of female and male speech

From Figure 7, we can find that the two spectrograms are quite similar, which make it hard to design a algorithm classifying them with high correctness simply based on spectrograms. While in Figure 8, the diffrence between cepstrogram of female and male speech is much more distinguishable, with different color distribution in each cepstral coefficient series. For a computer, it is feasible to discover that the two phrases are identical through the cepstrograms instead of the spectrograms.

4.3 Correlation matrices comparison of female speech

In this section, we plot and compare the correlation matrices for the logged spectral and normalized cepstral coefficient series of female speech. The result is shown in Figure 9 and Figure 10.

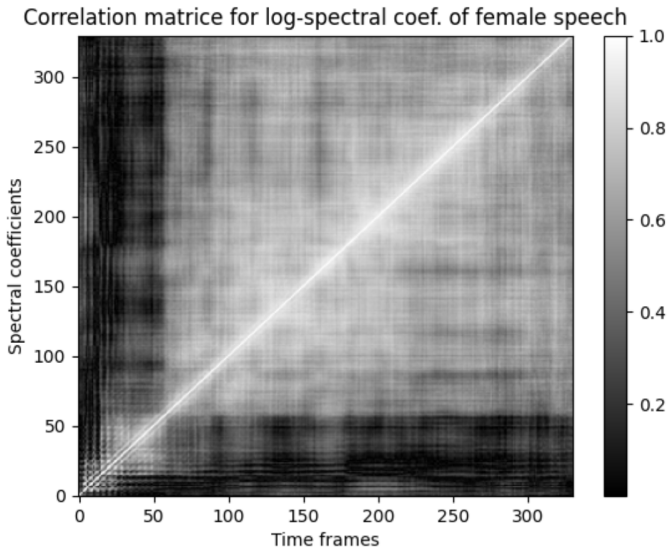


Figure 9: The correlation matrices for spectral coefficients of female speech

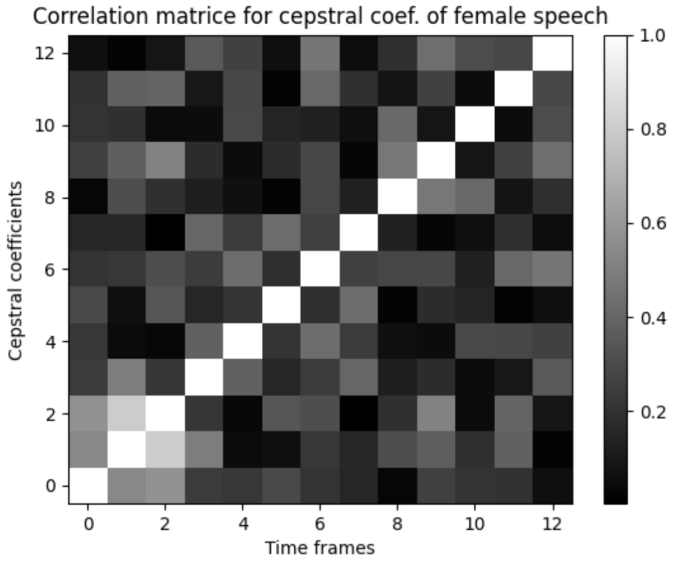


Figure 10: The correlation matrices for cepstral coefficients of female speech

From two figures we can conclude that neighbouring frequencies in the short-time spectrum can be highly correlated represented by large bright region in the figure. While MFCCs have smaller correlation between coefficients, with most bright area gathered in the diagonal of the matrice. This means that the features of the sound signal are better extracted in cepstrograms, which also accounts for the conclusion about the computer recognition we got in Section 4.2. Furthermore, the cepstrogram discards the high frequency information of the signal. It is possible to get similar MFCCs when people hear different sounds, like a male and a female say a same phrase, but in different pitches. It is also possible that people hear similar sounds with substantially different MFCCs if the sounds consist mainly of high frequency components, like sounds played by violin.

5 Conclusion

Both DFT and MFCC can perform the extraction of signal features. MFCC encodes the physical information (spectral envelope and details) to obtain the feature vector of the speech signal, which can be simply interpreted as the distribution of energy of the speech signal across different frequency ranges. The cepstrogram is superior to the spectrogram for computer recognition as it can achieve better decorrelation.