EN2202 Assignment 2

Sam Lewis   Bjoern Fischer
svrlewis@kth.se

1   Plots of the sound signals over the time domain

The female speech signal was plotted and sections of the signal were examined to find the oscillitory behavior of voiced segments and the noisy behaviour of unvoiced segments. The code that was used to create these plots can be found in `timePlots.m`.
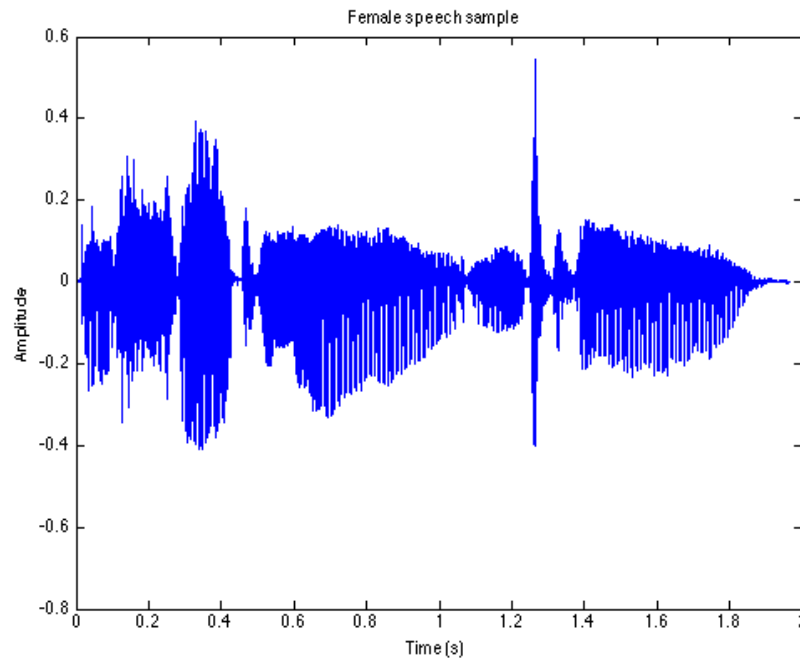

Figure 1: The complete female speech signal

The voiced sections of the signal were able to be located by looking for sections of the system that were oscillitory in nature - with some sort of repitition.
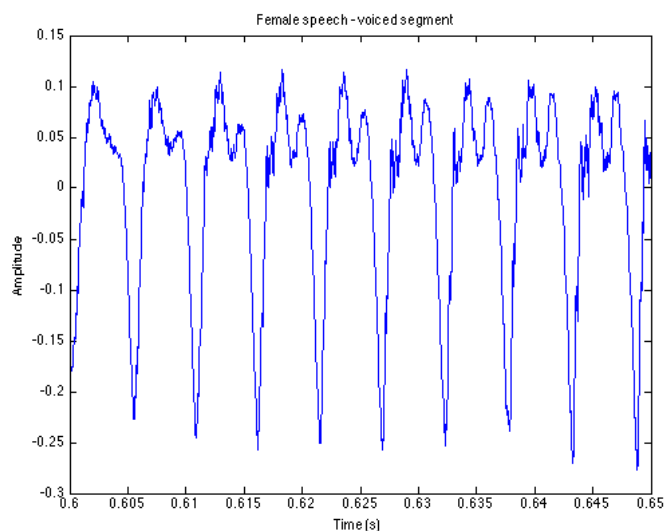

Figure 2: A voiced section of the female speech signal

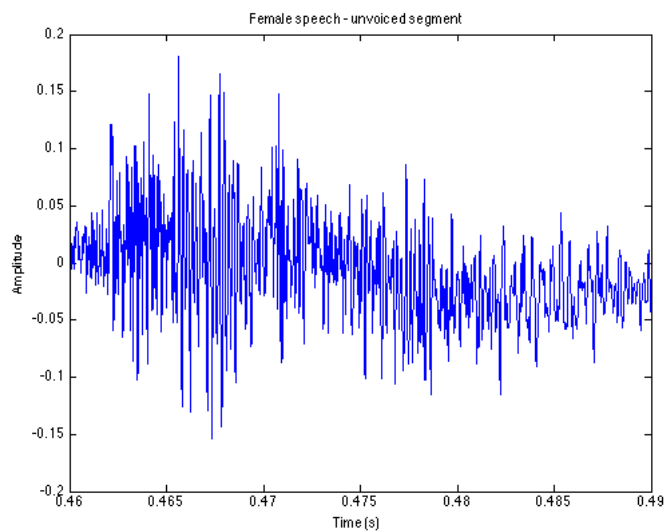The unvoiced sections were then found by looking for noisy sections without any sort of pattern.

Figure 3: An unvoiced section of the female speech signal

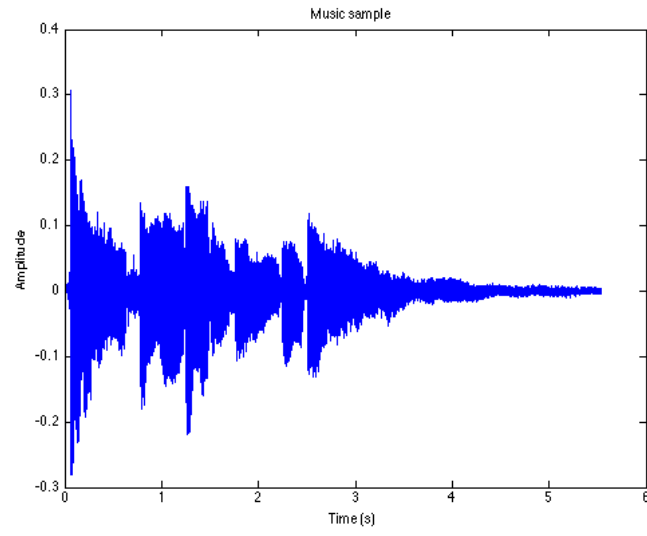Finally, the music signal was also plotted.

Figure 4: The complete music signal

2   Spectrograms

Spectrograms for the two signals shown previously were also plotted. For the music signal, harmonics representing the music sample were highlighted.
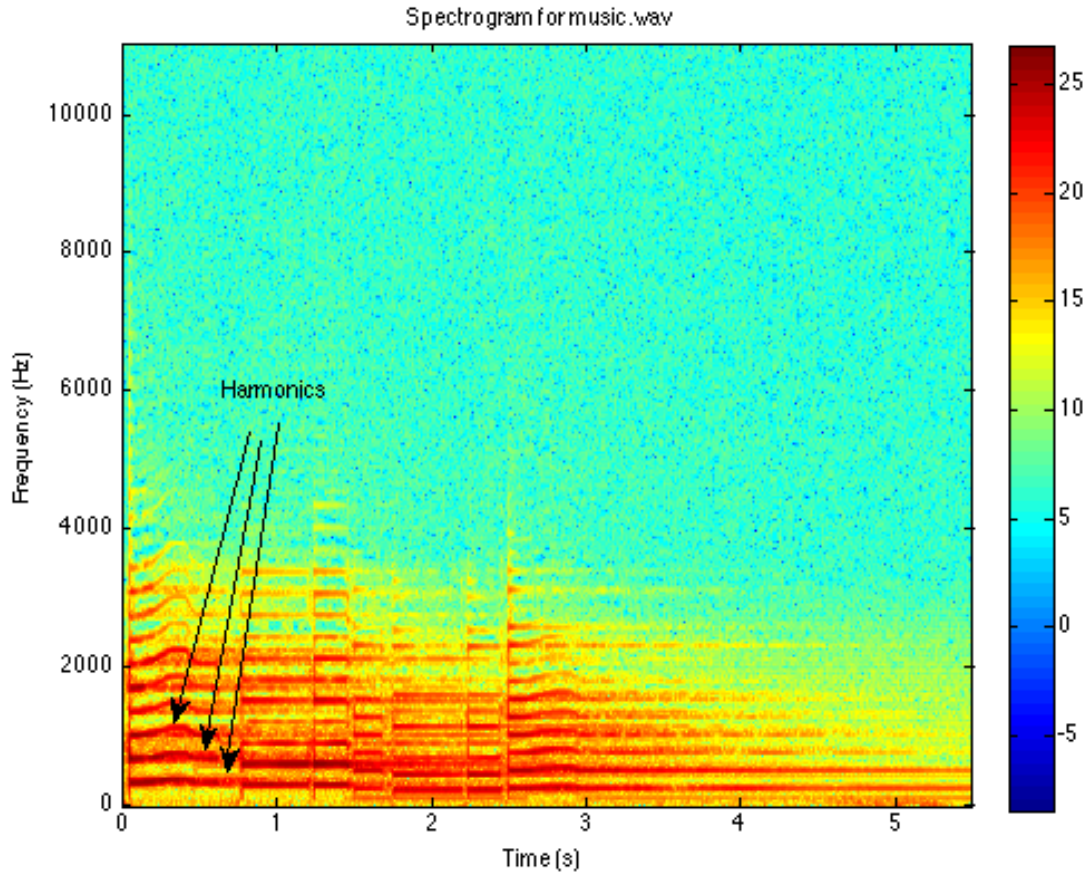

Figure 5: Spectrogram for the music speech signal

For the female speech sample, it was possible to locate the segments of voice by looking for areas in the spectrogram where there were harmonics. On the other hand, segments of unvoice were obvious from the areas in the spectrogram where there was no pattern in the frequency distribution. One voice segment and one unvoiced segment were highlighted on the diagram as seen below.
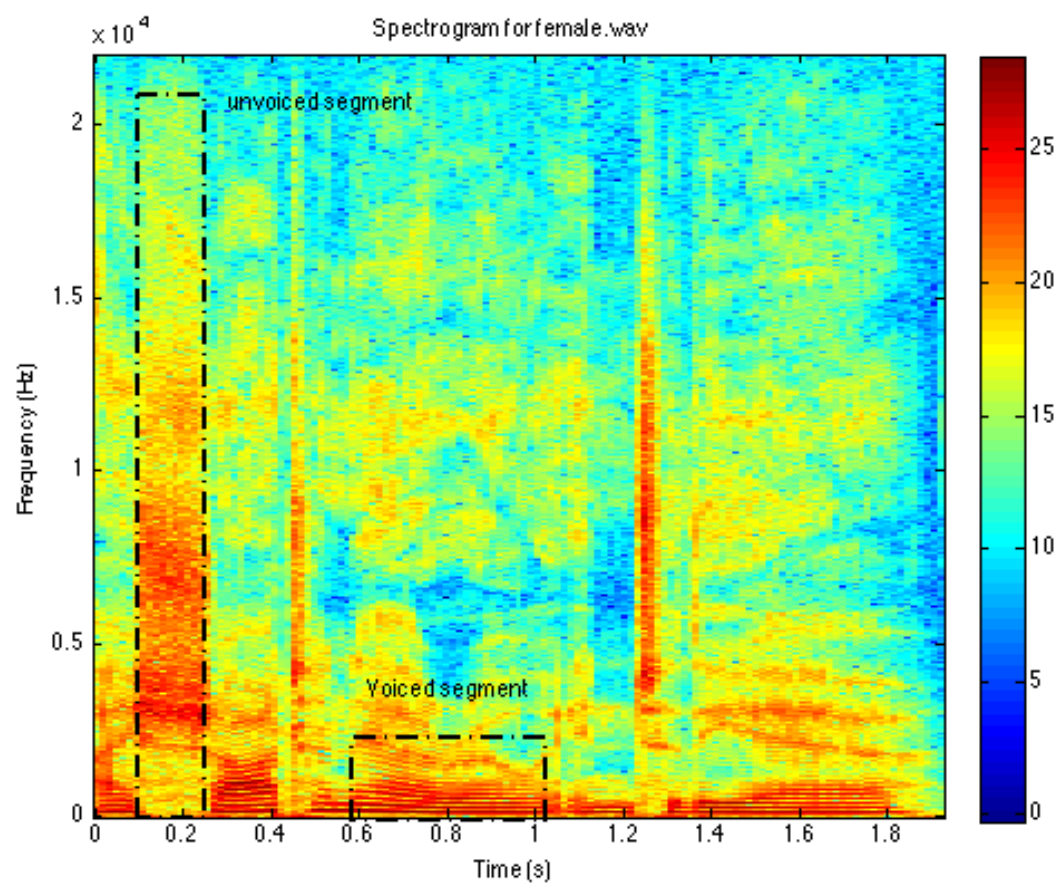
Figure 6: Spectrogram for the music speech signal

3 Cepstrograms

Normalised cepstrograms were also found and then plotted, alongside the spectrograms for the three different signals. The figure below shows the spectrogram and cepstrogram for the music file.
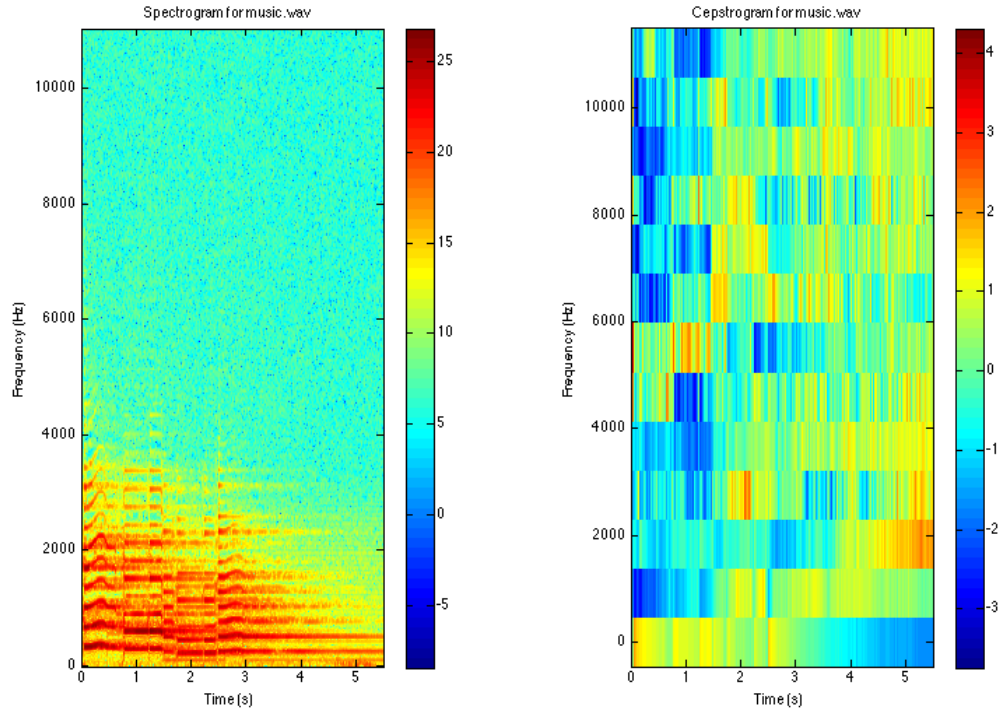

Figure 7: Spectrogram and cepstrogram for the music speech signal

Then, to analyse the differences between a male and female speaking the same phrase, the spectrogram and cepstrogram of the male and female speech samples were plotted together. The script to create these plots is `cepstrograms.m`.
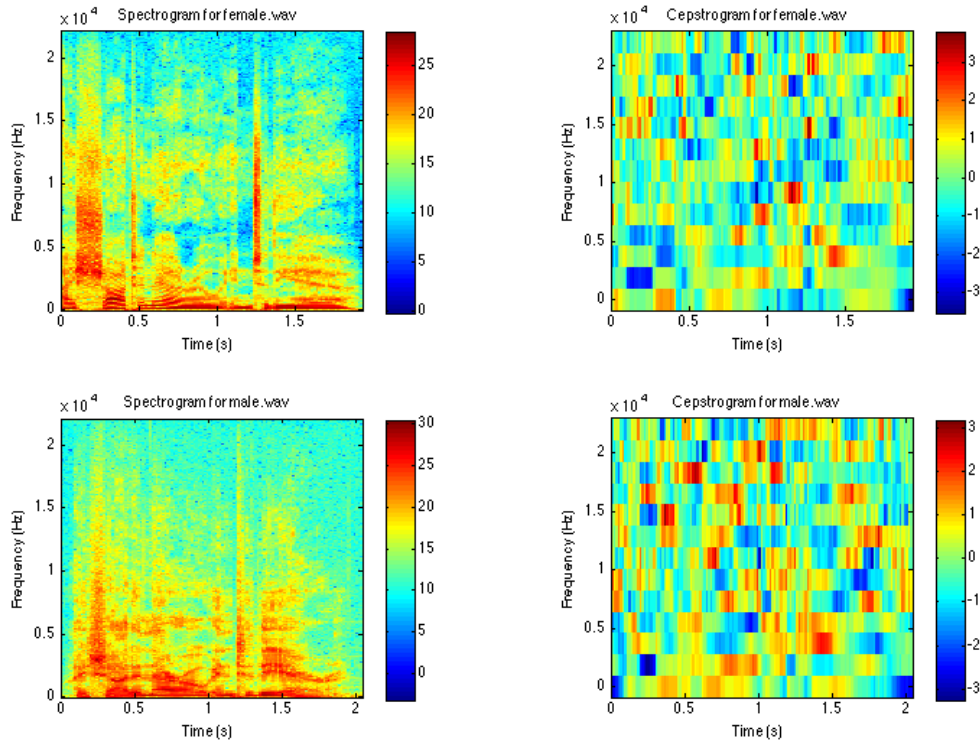

Figure 8: Spectrogram and cepstogram for the male and female speech signals

4   Correlation matrices

For the female speech sample, the correlation matrices for the spectral and cepstral coeffecient series were found. The MATLAB code used to create these plots is in `correlation.m`.
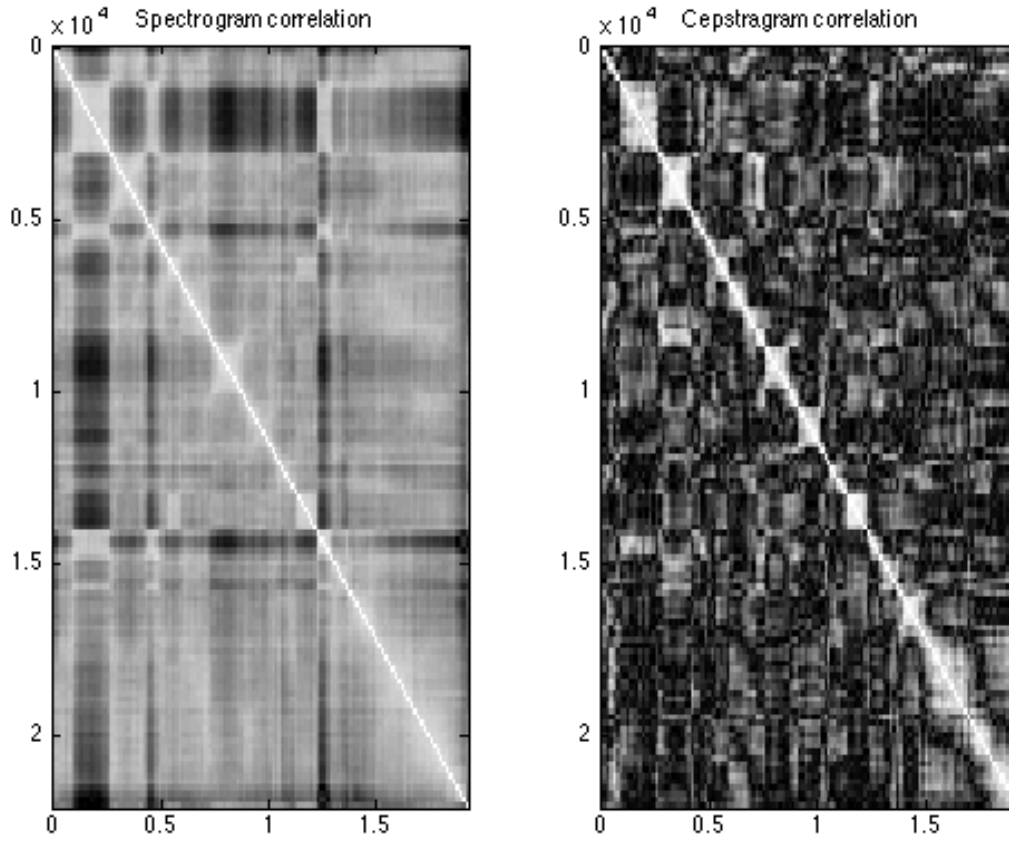

Figure 9: Spectrogram and cepstogram for the male and female speech signals

5   Questions

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

> Personally, for me, the spectrogram is much easier to interpret. As a human it is easier to filter the information and have an idea about the general trend of the signal so that, for example, it is possible to interpret when there are sections of voice or unvoice. The spectrogram is not as easy to understand and looks a lot more unorganised.

Can you see that they represent the same phrase? Could a computer discover this?

> Comparing the male and female spectrograms for the same phrase, it is possible to find the similarities and see that certain sounds, though they may be said with slight differences because of the gender and individual accents of the speaker, are in the same order.

> However, doing this sort of recognition would be hard for a computer as there is no mathematical rigour in comparing the two septrograms. More processing would need to be done.

Can you see that the cepstrograms represent the same phrase? What about a computer?

> Comparing cepstrograms is much harder for me as it is harder to spot patterns. However, presenting the information in this way might make it easier for a computer to view how similar the signals are. This is because the information is more clearly broken up into chunks.

Which matrix, spectral or cepstral, looks the most diagonal?

> Looking at the plots in figure 9, the cepstrogram looks the most diagonal. This is as the areas outside the diagonal are darker than in the correlation of the spectrogram, meaning the intensity outside of the diagonal is weaker.

How could a MFCC represenation in a speech recognizer be confused?

> To improve the performance in pattern recognition for speech recognition, the pitch of the signal is removed. Although pitch is not important in the recognition of individual words, it still carries important information in complete phrases. For example, the phrase 'you are an engineer' in English can be both a question or a statement depending on the intonation. Because the pitch of the words in the phrase are discarded in the speech recognition system, it would not be possible to find the meaning of the phrase.