Pattern Recognition Assignment 2

Vamshi Krishna Muga (muga@kth.se), Varsha Kirani Gopinath (varsha@kth.se)

October 11, 2015

Abstract

In this project we aim to develop a speech recognition system . In the second assignment, we get familiarized with the idea of Mel frequency Cepstral Coefficients(MFCC) and the way they are used in a speech recognition system. We solve various tasks in matlab and report the results in this paper.

1 Task 1: Code and plots of female speech and the music signal over time

Copy of getPlot code have been attached along with this report and the same can be found in appendix section.

Figure.1 represents the plots of female and music signals over time. Figures.2,3 represents the plots zoomed to 20ms which illustrate the oscillatory behavior.

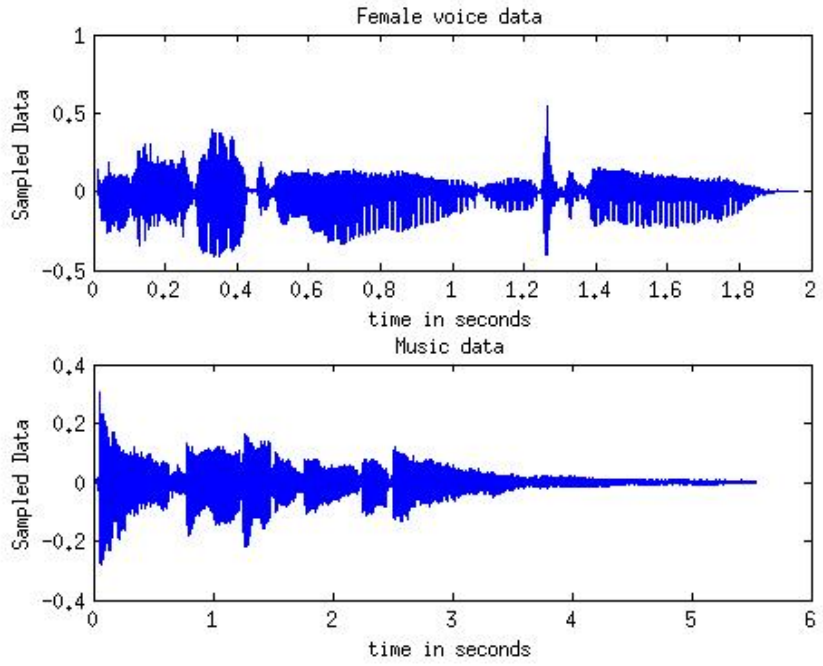

Figure 1: Plots of female speech and the music signal over time

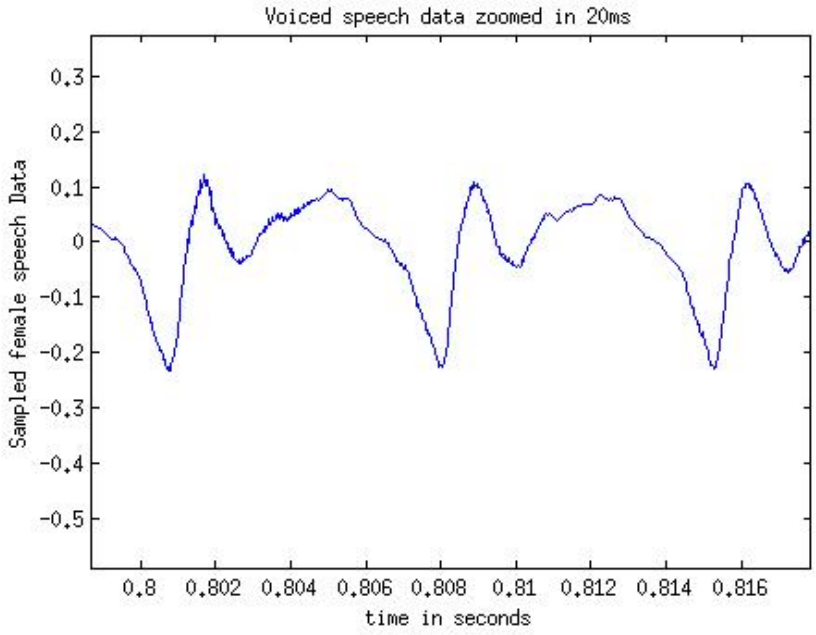

Figure 2: Zoomed in plot of voiced speech segment over 20ms time

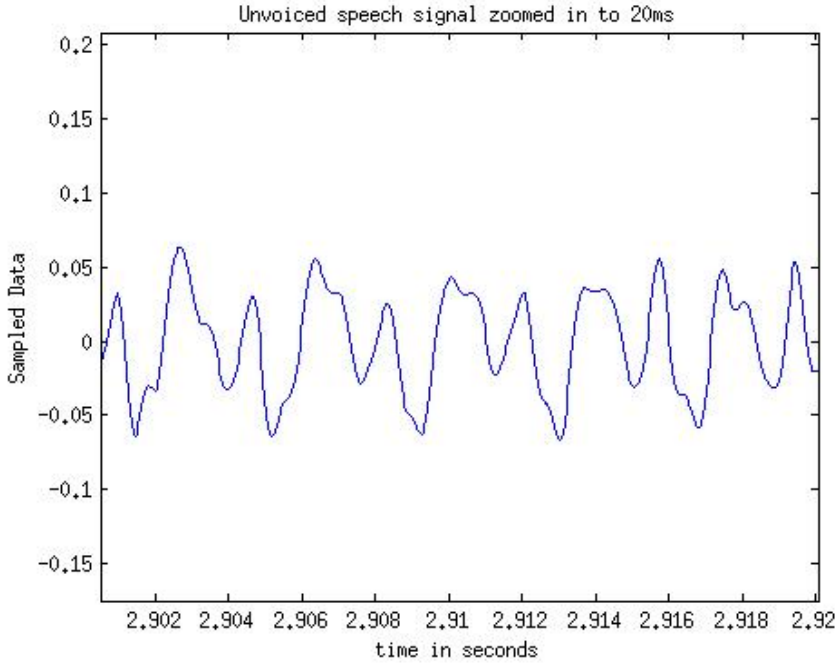

Figure 3: Zoomed in plot of unvoiced speech segment over 20ms time

2 Task 2: Code and plots of spectrograms for the same two signals.

Copy of spectgramPlots.m code has been attached along with this report and the same can be found in appendix section.

In figure.4, the sepctrogram of music sample is plotted and we can also notice the occurrence of harmonics in it. In figure.5, the spectrogram of female voice sample is plotted and it contained both voiced and unvoiced segments. It is difficult to recognize the harmonics here.

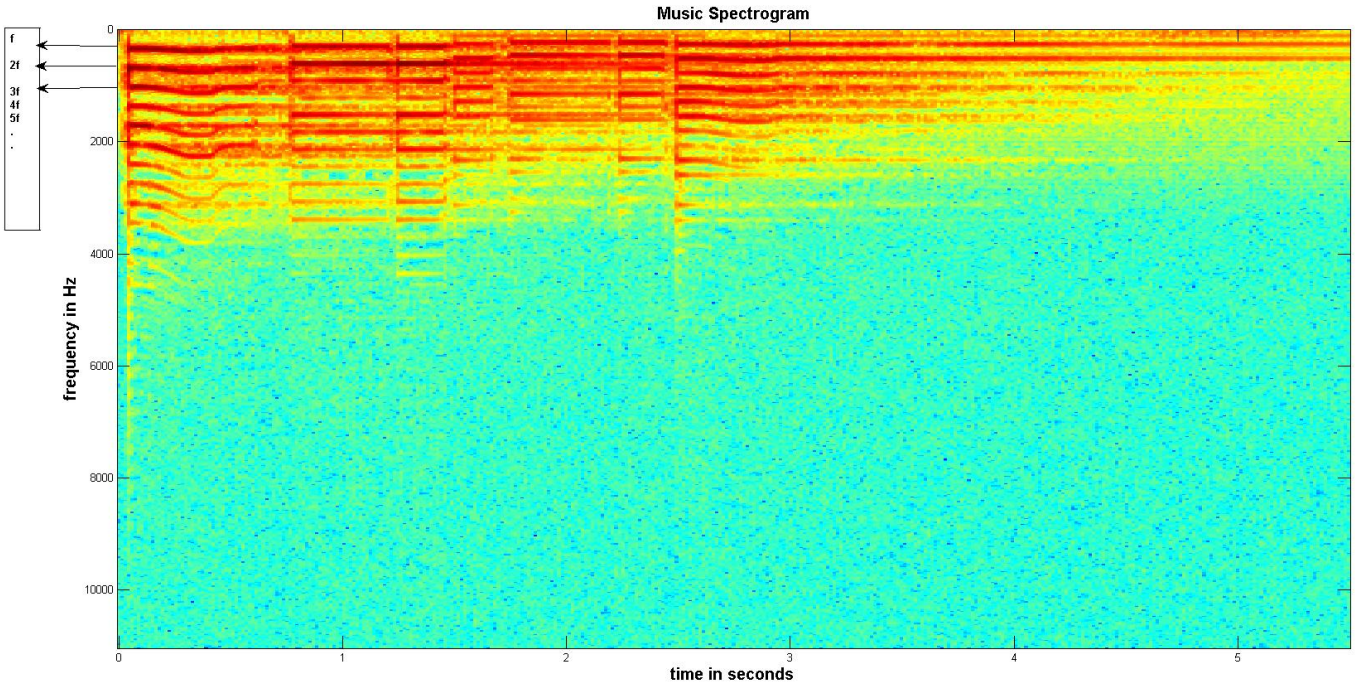

Figure 4: Spectrogram of music sample with harmonics being annotated

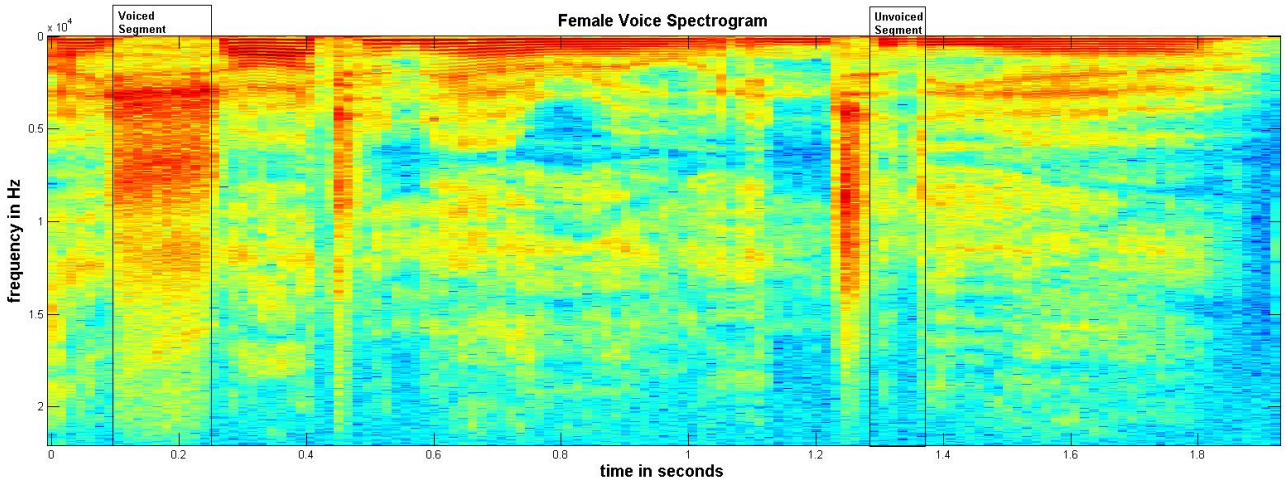

Figure 5: Spectrogram of female voice sample along with unvoiced segments.

3 Task 3: Comparison of Cepstrogram and Spectrogram representations for the same phrase

In figure.6, spectrogram representations of male and female voice are plotted for the same phrase. It can be seen that there is a similar pattern in the coefficients time series.

In figure.7, the first two cepstral coefficients are plotted for both the female and male voice. As seen, the variations in the coefficient time series are similar for both the voices.

In figure.8, five cepstral coefficients are plotted.

Matlab code for the same can be seen in appendix in cepSpec.m chapter and also attached along with this report.

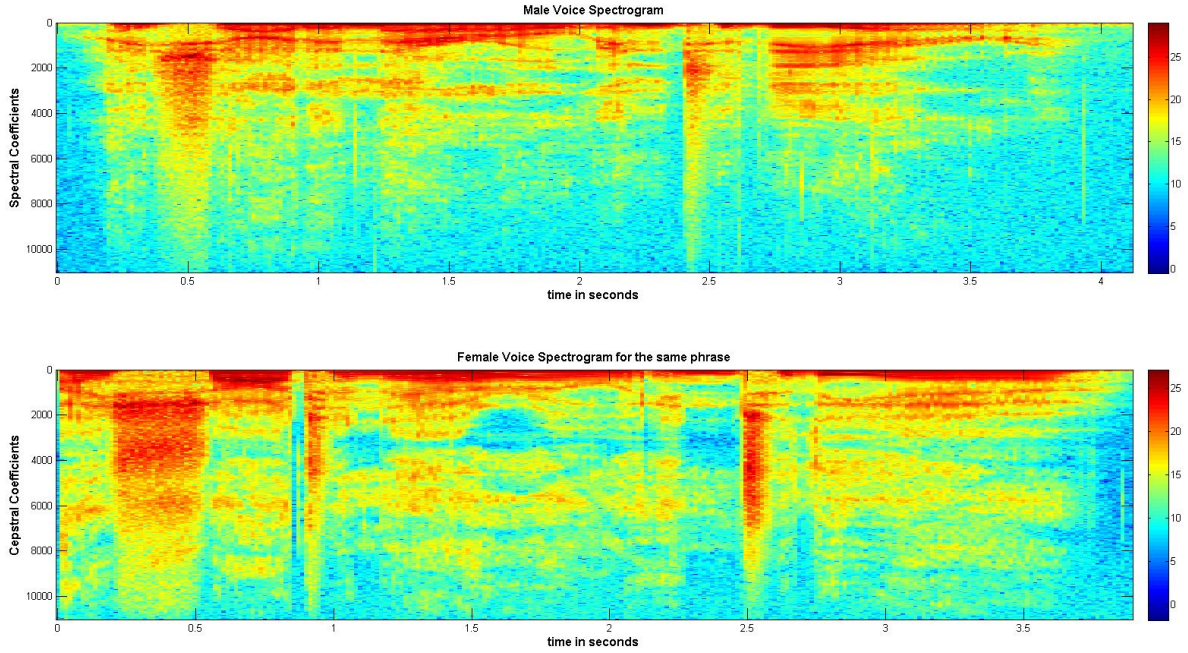

Figure 6: Comparison of Spectrogram representation of male and female voice for the same phrase

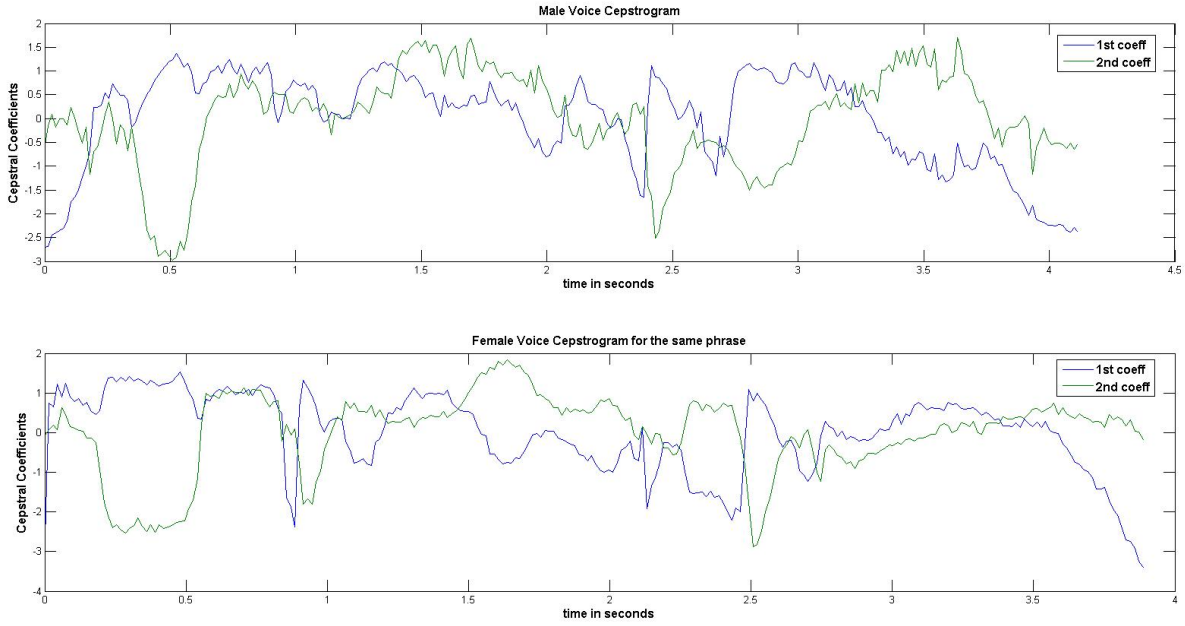

Figure 7: Comparison of Cepstrogram representation of male and female voice for the same phrase

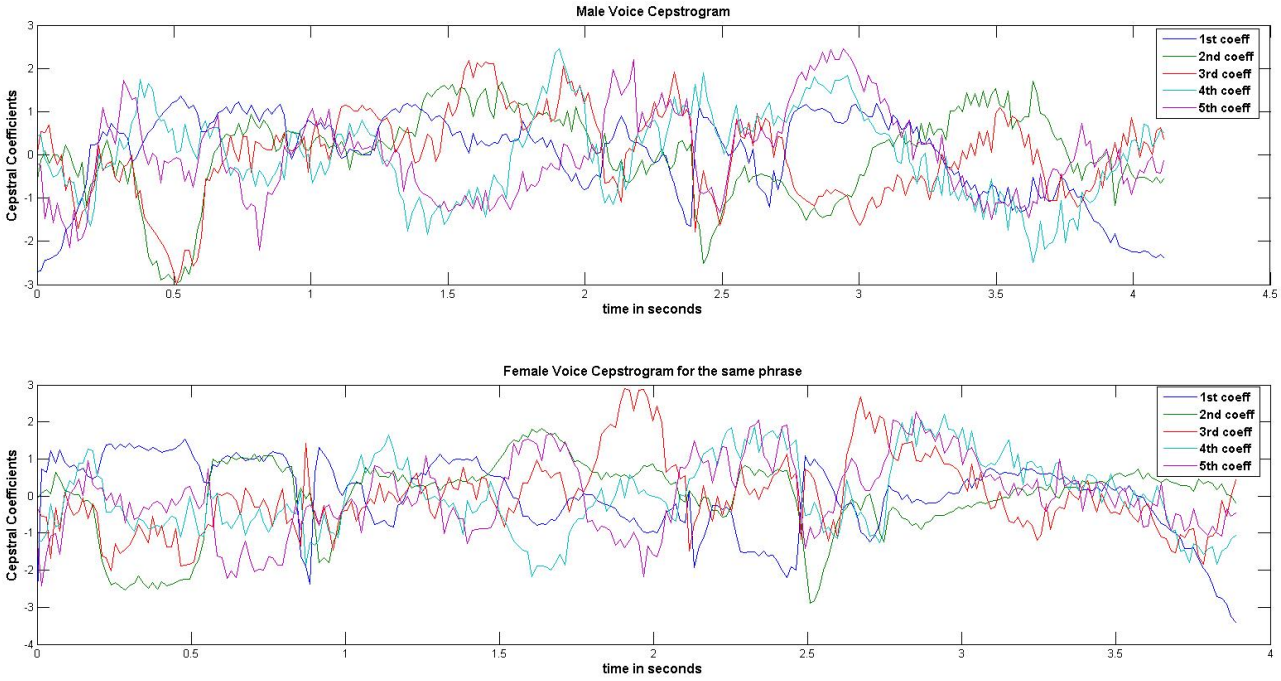

Figure 8: Comparison of five Cepstrogram coefficients of male and female voice for the same phrase

4 Task 4: Comparison of correlation matrices of spectral and cepstral coefficient time series

Fig.9 represents the correlation matrix of Spectral coefficient time series and fig.10 represents the correlation matrix of Spectral coefficient time series. As seen in the plots, the cepstral coefficient correlation matrix has a much more darker region around the diagonal than its counterpart in Fig.9. This indicates that the spectral coefficients in fig.9 are correlated to a greater extent.

Matlab code for the same can be seen in appendix in CorrelationCalc.m chapter and also attached along with this report.

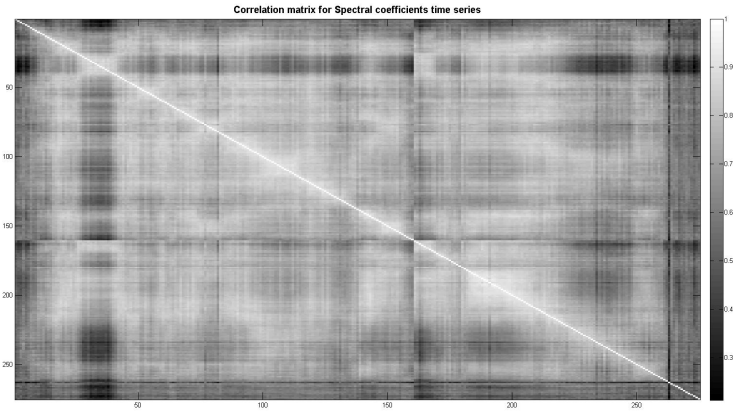

Figure 9: Correlation matrix of Spectral coefficient time series

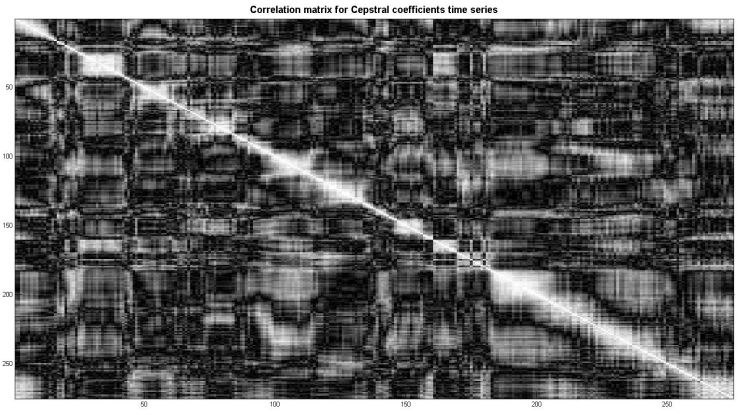

Figure 10: Correlation matrix of Cepstral coefficient time series

5 Task 5: Answers to the questions in the text associated with the plots

5.1 Q1: Which representation(Spectrogram or Cepstrogram) is easier to interpret for a computer and for a human?

Figure.11 represents the spectrogram and cepstrogram of female and music sample. From the figure we can see that the spectrogram is visually better to interpret for humans. However, a cepstrogram can capture the same information in less number of coefficients. Hence, for a computer it might be more easy to use the cepstrogram for easier numerical computation.

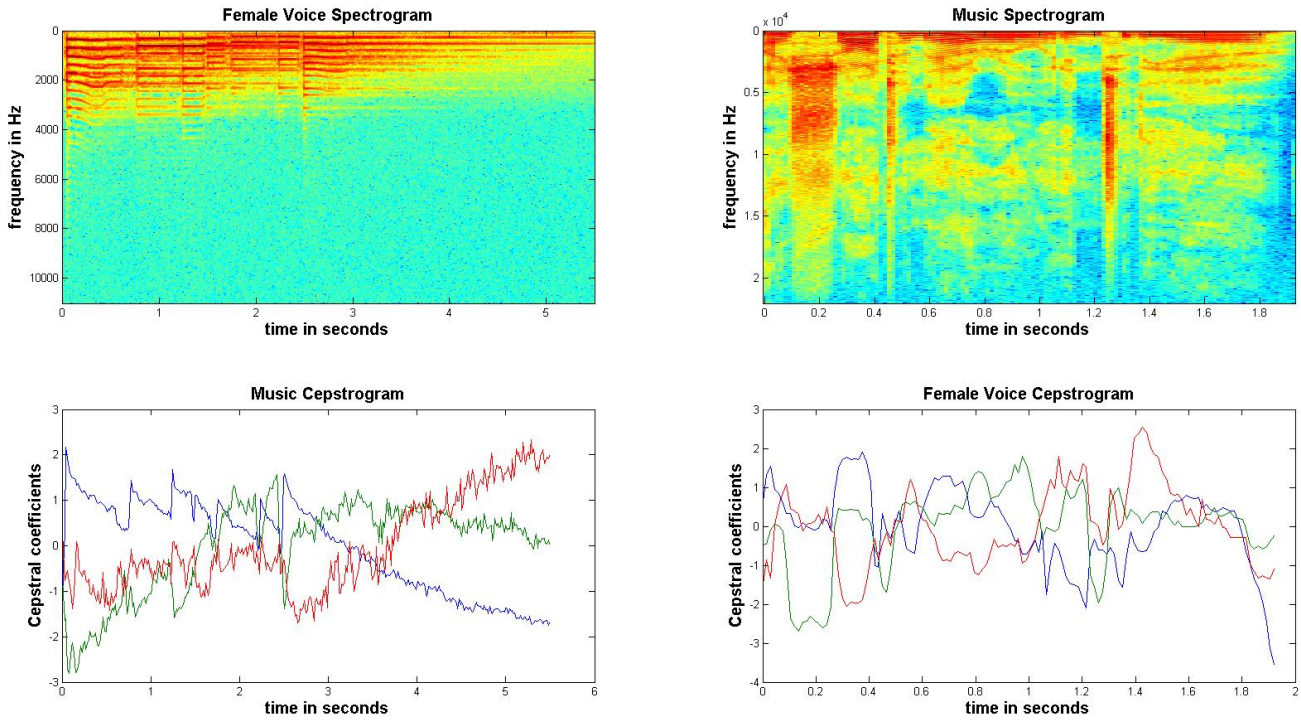

Figure 11: Comparison of spectrograms and cepstrograms of female and music sample

5.2 Q2: Can a computer decide if two phrases are same?i)given cepstrogram ii)given spectrogram

By observing figure.6, a computer might not be able to decipher if both the voices carry the same phrase. This is because the dimension of the feature vector is high.

However for figure.7, the feature vector is of less dimension(typically 13) and it is easier for numerical calculations.

6 Task 6: Normalized static and dynamic mfccs

Copy of matlab function to get dynamic mfccs have been attached along with this report and the same can be found in appendix section(getDynamicFeatures.m).

7 Task 7:

In MFCCs computation, the last step involves taking a discrete cosine transform. This step is mainly done to decorrelate the inputs and reduce the dimensionality. Also, this step is not directly related to human sound perception unlike the other ones. Hence, we do loose some information.

7.1 Example for same MFCC with different human perception :

We can think of a situation where the same note played with different instruments sounds different for humans. However, the frequency content being similar gives the same mfccs.

7.2 Example for different MFCC with similar human perception :

When two similar sound clips are played in very noisy circumstance, human ears can still match or hear both the sounds to be similar. However, mfccs are not so resistant to background noise and hence they may vary significantly.

Appendices

getPlots.m

```
close all;
clear all;
clc;

%reading all the sound files
%Fs is sample rate and y is data

hFile1 = 'female.wav';
[yFemale,Fs1] = wavread(hFile1);

hFile2 = 'music.wav';
[yMusic,Fs2] = wavread(hFile2);

hFile3 = 'male.wav';
[yMale,Fs3] = wavread(hFile3);

ts1 = 1/Fs1;
ts2 = 1/Fs2;
ts3 = 1/Fs3;
```

```
%used in plotting for the time scale
t1 = 0:ts1:(length(yFemale) * ts1 ) - ts1;
t2 = 0:ts2:(length(yMusic) * ts2 ) - ts2;
t3 = 0:ts3:(length(yMusic) * ts3 ) - ts3;

%Plot of Female voice sample
subplot(2,1,1);
plot(t1,yFemale);
xlabel('time in seconds');ylabel('Sampled Data');title('Female voice data');


%Plot of Music sample
subplot(2,1,2);
plot(t2,yMusic);
xlabel('time in seconds');ylabel('Sampled Data');title('Music data');
```

spectgramPlots.m

```
close all;
clear all;
clc;

%reading sound file
% y denotes data and Fs denotes sample data
hFile1 = 'female.wav';
[yFemale,Fs1] = wavread(hFile1);

hFile2 = 'music.wav';
[yMusic,Fs2] = wavread(hFile2);

%Music Spectrogram
winlength = 0.03;
[spectgram,f,t] = GetSpeechFeatures(yMusic,Fs2,winlength);
%logarithm is used for numerical tractability
imagesc(t,f,log(spectgram));
xlabel('time in seconds');
ylabel('frequency in Hz');
title('Music Spectrogram');
%making bold texts
set(findall(gcf,'type','text'),'FontSize',15,'fontWeight','bold');

figure;

%Female voice spectogram
winlength2 = 0.03;
[spectgram2,f2,t22] = GetSpeechFeatures(yFemale,Fs1,winlength2);
%logarithm is used for numerical tractability
imagesc(t22,f2,log(spectgram2));
```

```
xlabel('time in seconds');
ylabel('frequency in Hz');
title('Female Voice Spectrogram');
%making bold texts
set(findall(gcf,'type','text'),'FontSize',15,'fontWeight','bold');
```

cepSpec.m

```
close all;
clear all;
clc;
hFile1 = 'female.wav';
[yFemale,Fs1] = wavread(hFile1);

hFile3 = 'male.wav';
[yMale,Fs3] = wavread(hFile3);
winlength = 0.03;
[mfccs_m,spectgram_m,f_m,t_m]=GetSpeechFeatures(yMale,Fs3,winlength,13);
[mfccs_f,spectgram_f,f_f,t_f]=GetSpeechFeatures(yFemale,Fs1,winlength,13);

mfccs_m = MFCC_Norm(mfccs_m);

mfccs_f = MFCC_Norm(mfccs_f);

subplot(2,1,1);imagesc(t_m,f_m,log(spectgram_m));
xlabel('time in seconds');
ylabel('Spectral Coefficients');
title('Male Voice Spectrogram');
subplot(2,1,2);imagesc(t_f,f_f,log(spectgram_f));
xlabel('time in seconds');
ylabel('Cepstral Coefficients');
title('Female Voice Spectrogram for the same phrase');
 %making bold texts
set(findall(gcf,'type','text'),'FontSize',12,'fontWeight','bold');
figure;
subplot(2,1,1);plot(t_m,mfccs_m(1:5,:));
xlabel('time in seconds');
ylabel('Cepstral Coefficients');
title('Male Voice Cepstrogram');
subplot(2,1,2);plot(t_f,mfccs_f(1:5,:));
xlabel('time in seconds');
ylabel('Cepstral Coefficients');
title('Female Voice Cepstrogram for the same phrase');
%making bold texts
set(findall(gcf,'type','text'),'FontSize',12,'fontWeight','bold');
```

MFCCNorm.m

```
function [ mfcc ] = MFCC_Norm( mfcc )
```

```
        [m,n] = size(mfcc);
        mfcc_old = mfcc;
        %making zero mean
        for i=1:m
        mfcc(i,:) = (mfcc(i,:) - mean(mfcc_old(i,:)));
        end

        mfcc_old2 = mfcc;
        %making unit variance
        for i=1:m
            mfcc(i,:) = mfcc(i,:)/sqrt(var(mfcc_old2(i,:)));
        end
end
```

CorrelationCalc.m

```
close all;
clear all;
clc;
%frame length
winlength = 0.03;

%reading male voice sound clip
hFile3 = 'male.wav';
[yMale,Fs3] = wavread(hFile3);

%feature extraction
[mfccs_m,spectgram_m,f_m,t_m]=GetSpeechFeatures(yMale,Fs3,winlength,13);
mfccs_m = MFCC_Norm(mfccs_m);
figure;

%logarithm used for numerical tractability
corr_spect = corr(log(spectgram_m));
corr_cept = corr(mfccs_m);

%gray scale to make easier analysis
colormap gray;
%absolute to remove negatives
imagesc(abs(corr_spect));
title('Correlation matrix for Spectral coefficients time series');
colormap gray;
%making bold texts
set(findall(gcf,'type','text'),'FontSize',15,'fontWeight','bold');
figure;

imagesc(abs(corr_cept));
title('Correlation matrix for Cepstral coefficients time series');
colormap gray;
```

```
%making bold texts
set(findall(gcf,'type','text'),'FontSize',15,'fontWeight','bold');
```

getDynamicFeatures.m

```
function [ deltas,delta_deltas ] = getDynamicFeatures( mfcc )

        %first derivative along columns (each frame)
        deltas = diff(mfcc,1,2);

        %second derivative along columns (each frame)
        delta_deltas = diff(mfcc,2,2);

end
```