CS-215: Experiment 7B

Soumen Pradhan – 1912176

13 . 04 . 2021

1. Identification of vowels in Audio Sample

Aim

To perform Spectral Decomposition on an audio sample and identify the vowels occuring in it.

Theoretical Background

Discrete FT

The discrete Fourier transform which transforms a sequence of N complex numbers $\{x_n\} \coloneqq x_0, x_1, \ldots, x_{N-1}$ into another sequence of complex numbers $\{X_k\} \coloneqq X_0, X_1, \ldots, X_{N-1}$ is defined by -

$$X_k = \sum_{n=0}^{N-1} x_n \cdot \exp -\frac{j2\pi}{N} kn$$

Usually the x_n are samples of a signal and the DFT shows the frequency density of the signal.

Formants

Formants are defined as broad peaks or local maximums in the frequency spectrums. Humans, despite having diffrent vocal tract lengths (and hence different resonant frequencies), can produce sounds perceived to be in the same category.

To solve this solution, Hermann coined the term 'formant'. A vowel can be uniquely defined by a series of formant frequency (within acceptable variance). [1]

Methodology

- The first 100 samples of handel.mat are plotted.
- The fourier transform and spectrogram of the samples are computed and plotted.

[1] McKendrick, J. G. (1903). Experimental phonetics. In Annual report of the board of regents of the Smithsonian institution for the year ending June 30, 1902 (pp. 241–259). Smithsonian Institution.

- The coefficients are calculated for a small interval using LPC function and are then solved to find complex roots
- The frequency and the corresponding bandwidth are calculated and sorted from the complex roots (only those having positive imaginary parts).
- Then, the triplet values F_1, F_2, F_3, are compared against different vowel formant data.

Code

```
clearvars
clc
load handel
y = y';

% Isolating the samples 21k - 26k. Plot 1st 100.

vow = y(21000:26000);
time = (0: length(vow)-1) / Fs;

ogPlt = axes;
stem(time(1:100), vow(1:100), 'k' ,'Marker', 'None');
ogPlt.PlotBoxAspectRatio = [2 1 1];
axis([-time(10) time(110) -.8 .8]);
grid on

xlabel('Time (s)', 'FontSize', 8);
ylabel('Amplitude', 'FontSize', 8);
print('og_signal.eps', '-depsc');

% Plot the Fourier Transform and Spectrogram.

ft = fft(vow);
N = length(vow);
freq_shift = (-N/2:N/2 - 1) * (Fs/N);

ftPlt = subplot(2, 1, 1);
stem(freq_shift, abs(fftshift(ft)), 'k', 'Marker','none');
grid on

ftPlt.XLim = [0 freq_shift(end)];
ftPlt.Position = [0.1300 0.4512 0.67 0.4];
ftPlt.XTick = [];
ftPlt.YTick = 50:50:300;
ylabel(ftPlt, 'FT Amplitude', 'FontSize', 8);

spectPlt = subplot(2, 1, 2);
```

```
colormap('jet');
spectrogram(vow, 200, [], [], 'xaxis', Fs);

print('ft_spect.eps', '-depsc');

% Find the Formants for 0.1s intervals.

dt = 1/Fs;
formants = zeros(5, 3);

for i = [.1 .2 .3 .4 .5]
    I0 = round(i/dt);
    Iend = round((i + 0.1)/dt);

    testRange = vow(I0:Iend);
    testRange = testRange .* hamming(length(testRange))';
    testRange = filter(1, [1 .63], testRange);

    poly = lpc(testRange, 8);
    rts = roots(poly);
    rts = rts(imag(rts) >= 0);
    argz = atan2(imag(rts), real(rts));

    [frqs, indices] = sort(argz .* (Fs/(2*pi)));
    bw = -.5 * (Fs/(2*pi)) * log(abs(rts(indices)));

    nn = 1;
    for kk = 1:length(frqs)
        if (frqs(kk) > 90 && bw(kk) <400)
            formants(i*10, nn) = frqs(kk);
            nn = nn+1;
        end
    end
end

writematrix(formants, 'formants.txt');
```

Input Description

Samples considered from 'handel.mat' = 21e+3 - 26e+3

Spectrogram window = 200 samples

Interval for formant calculation = 0.1s

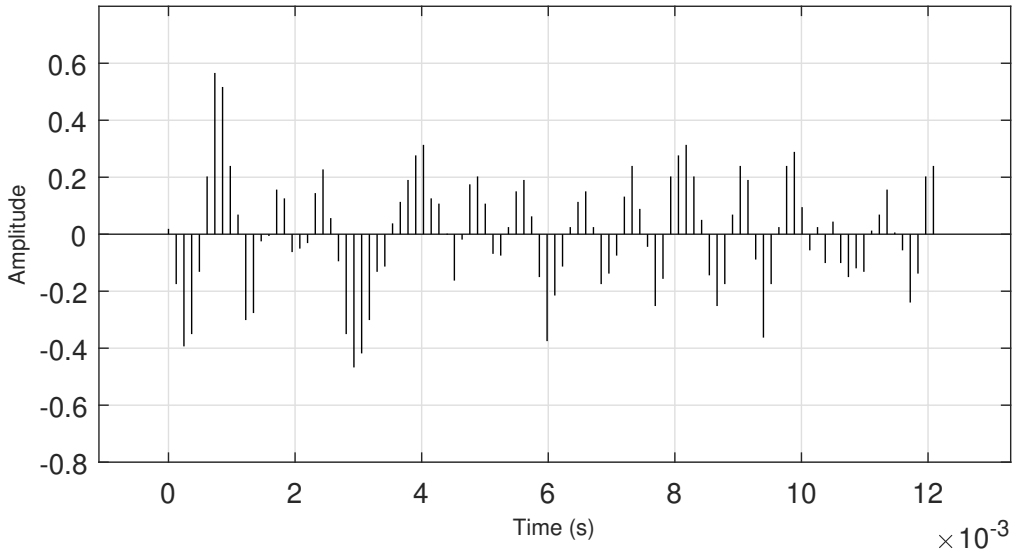

Figure 1.1: *First 100 samples of the audio (0.012s)*

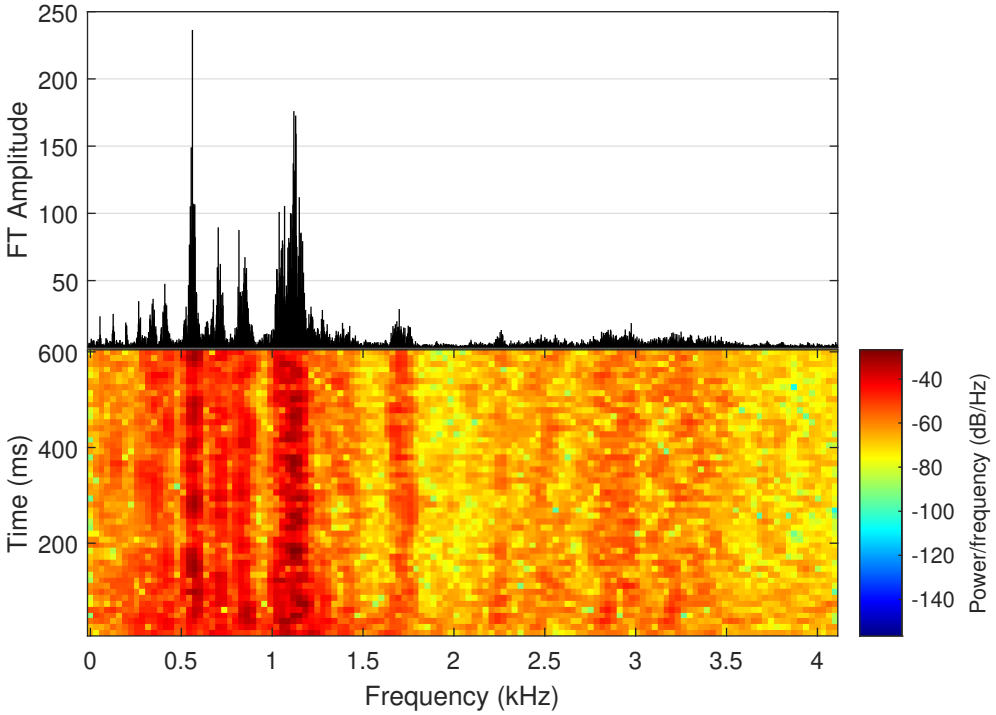

Figure 1.2: *Fourier Transform and the Spectrogram (Window of 200 sample)*

Result

As can seen, broad peaks are forming at lower frequencies while, the higher frequncies are rather sparse. Upon hearing the audio (length = 0.6s), the vowel ʌ is the closest guess.

Looking at the generated fricant results for 0.1s intervals, we get -

Interval (s)	$\boldsymbol{F_1}$ (Hz)	$\boldsymbol{F_2}$ (Hz)	$\boldsymbol{F_3}$ (Hz)
0.1 - 0.2	590.69	1149.43	2867.11
0.2 - 0.3	644.34	1151.44	2861.51
0.3 - 0.4	588.53	1133.52	2838.46
0.4 - 0.5	630.65	1156.74	2868.50
0.5 - 0.6	633.18	1139.03	2850.48

Table 1.1: *fffff*

Now, comparing with published formant values [2], we can clearly see all intervals have a /ʌ/ like sound. Here, the vowels /i/ /ɪ/ /ɛ/ /æ/ /ɑ/ /ɔ/ /ʊ/ /u / /ʌ/ /ɝ/ are considered.

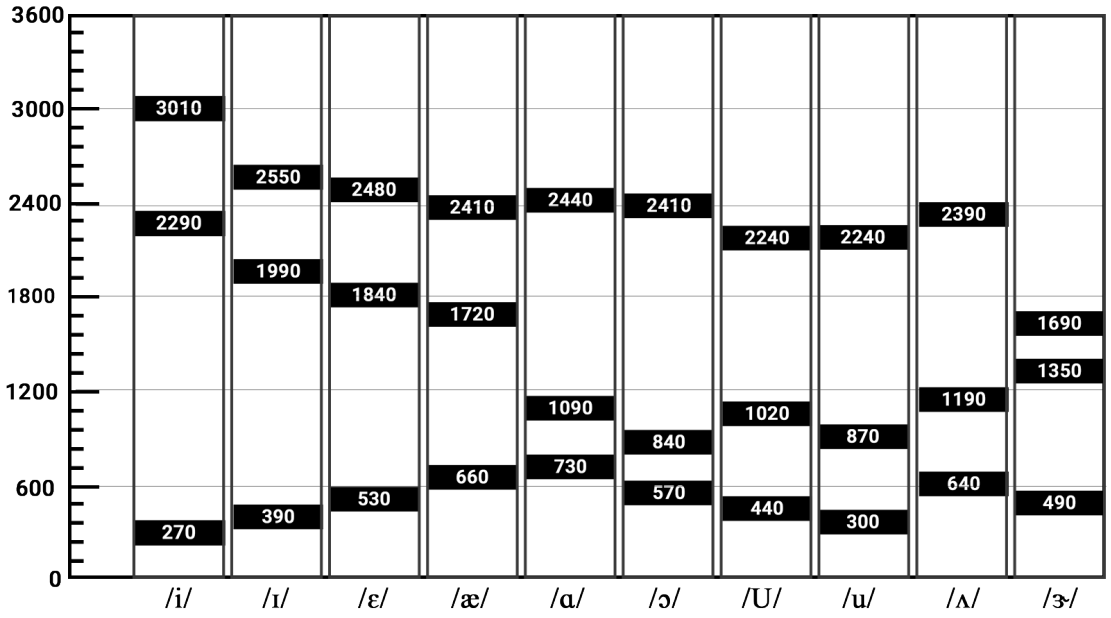

Figure 1.3: *Average formant locations (Hz) for vowels in American English*

Conclusion and Discussion

The vowel ʌ is clearly the most dominant sound in the sample. There is indeed a range of values centered around F_1 and F_2 frequencies.

Although, there is certainly a significant difference in values of F_3 frequency. This could be due to the fact that the subject is singing in the sample, while the published data is for normal speaking voice.

[2]Control Methods Used in a Study of the Vowels, Gordon E. Peterson and Harold L. Barney, The Journal of the Acoustical Society of America, Vol. 24, Pg. 175 (1952); doi: 10.1121/1.1906875