1 MATLAB Exercise on Formants

In this exercise, the linear predictor is used to predict the current sample of the speech signal using a linear combination of past samples, a filter of order 10 was chosen, thus 10 past samples was used for estimation. Also the frame size was 256 and the number of frames was 128 but with 50% overlapping of sample the estimation was looped for *256 data frames*.Location of the formants can be approximate by finding the roots of the LPC function. Since the filter has real valued coefficients the roots occur in conjugate pairs. the angle of the root on the pole zero plot gives the frequency of the formant.

Figure 1 is the plot of the first and second formants against the frame number, we can observe that there is a large difference between the frequencies for the first and second formants.

Table 1 contains the Formant 1 and 2 frequencies along with whether the frame is a voiced or unvoiced based on observation from the signal signatures From the signal signatures it is possible to distinguish if a signal is voiced or unvoiced. Voiced speech is quasi-periodic in the time-domain and harmonically structured in the frequency-domain. Unvoiced speech is random-like and broadband. *Figure 2* shows the LPC and the time domain plot of the unvoiced and voiced signal.

Frame #	Formant 1(Hz)	Formant 2(Hz)	Voiced or Unvoiced
50	402.90	1110	Voiced
51	407.31	1010	Voiced
52	435.89	958.5	Voiced
53	458.42	660	Voiced
54	696.76	1460	Unvoiced
55	713.81	1500	Unvoiced
56	756.12	1470	Unvoiced
57	753.49	1550	Unvoiced
58	732.14	1610	Unvoiced
59	760.62	1600	Unvoiced
60	581.79	689	Voiced
61	593.3	1290	Voiced
62	562.3	1481.0	Voiced
63	534.1	1464.8	Voiced
64	442.3	1411.5	Voiced
65	374.71	1360	Voiced
66	360	1430.9	Voiced
67	245.3	1372.3	Unvoiced
68	499.4	1255.3	Unvoiced
69	591.2	2203.3	Unvoiced
70	627.8	1790	Unvoiced
71	1202.6	1640.9	Unvoiced
72	984.42	1420	Unvoiced
73	497.29	1430	Unvoiced
74	404.6	1480	Voiced
75	448.9	1439.1	Voiced
76	487.82	1480.9	Voiced
77	484.31	1460.9	Voiced
78	484	1440	Voiced
79	367.7	1440	Unvoiced
80	346	1530.3	Voiced

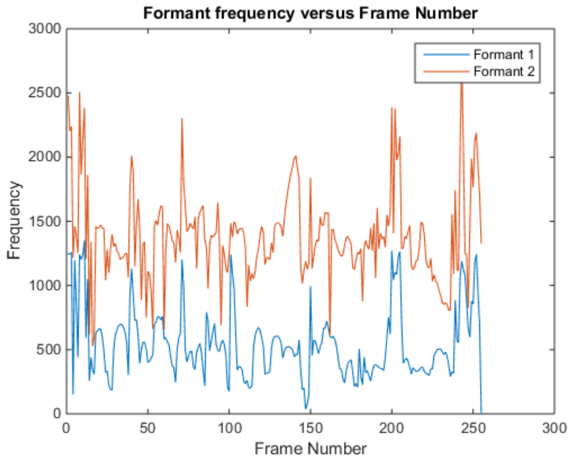
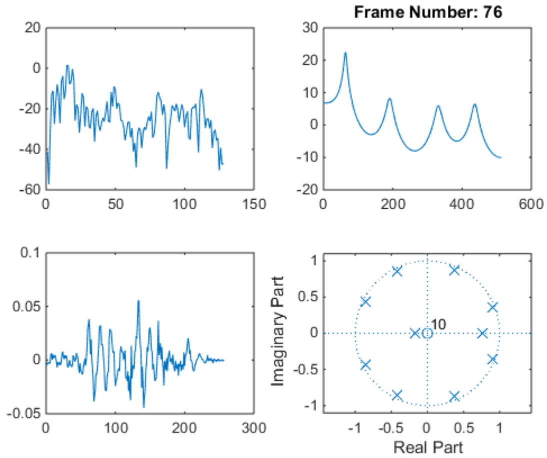
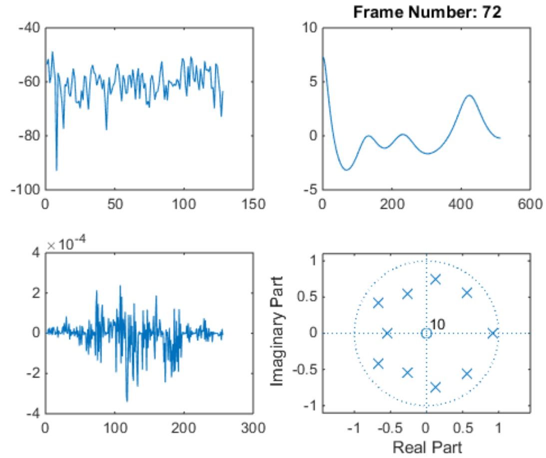


Figure 1: Plot of the first two formants across time (formants vs frame number) when number of frames(K=256)



(a) Voiced signal



(b) Unvoiced Signal.

Figure 2: FFT, LPC,Time Domain and Pole Zero Plot

2 J-DSP Exercise

K means algorithm is used for clustering the given speech formant datasets from 100 different speakers.The raw data is taken and centroids are assigned arbitrarily in the formant space. In every iteration, the distance metric between the data points and their respective centroids are calculated and the within cluster sum of squares is minimized. The new centroid is the mean of the points assigned to each centroid. This process is continued until convergence where the effective shift in the centroid is negligible and the mean square error is minimum. Here J-DSP software to implement the same. The centroid, mean square error and the required deviations are computed from it for further analysis

Values of the cluster centroids after convergence

For the Speech formant dataset 1 the value of the cluster centroids after convergence is tabulated below

Data Set	F1	F2
Vowel i	262.01	2403.02
Vowel u	240.39	592.81
Vowel aa	732.96	947.53
Vowel ae	754.21	1610.76

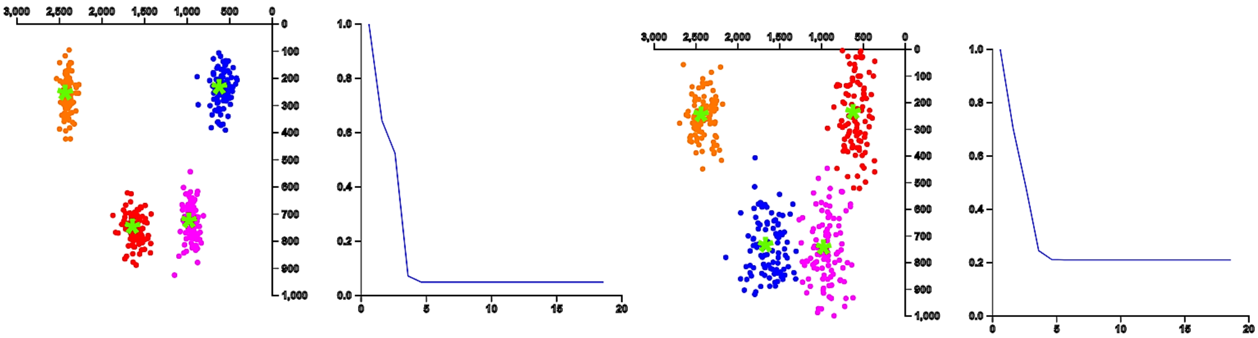
For the Speech formant dataset 2 the value of the cluster centroids after convergence is tabulated below

Data Set	F1	F2
Vowel i	253.63	2410.46
Vowel u	244.35	592.92
Vowel aa	751.98	944.52
Vowel ae	741.36	1637.38

Value of the within-cluster sum of squares observed after the convergence

For Data set 1: Final MSE: 0.0495, For Dataset 2: Final MSE: 0.2104

Clustering and Convergence curve



(a) For Speech Formant Dataset 1

(b) For Speech Formant Dataset 2

Figure 3: Clustering and Convergence curve

Deviation observed for each vowel's formant frequencies after convergence

Given that the average values of the formants is as tabulated below, the deviation from the centroid values that is obtained by clustering is as given below

Vowel	F1	F2	F1 Diff.	F2 Diff.	F1 Diff.	F2 Diff.
i	240	2400	-22.01	-3.02	-13.63	10.46
u	250	595	10.39	2.19	5.65	2.08
aa	750	940	17.04	-7.53	-1.98	-4.52
ae	750	1610	-4.21	-0.76	8.64	-27.38

Table containing the deviation of formant frequencies from the one obtained by clustering, the cells colored red denotes the evaluation for dataset 1, cells colored blue denotes evaluation for dataset2

Observations

As seen from the clustering plots for dataset 2 the values are a little more spread out compared to that i dataset1.
For *Vowel i*, in the case of Dataset 1, the F1 obtained by clustering is at a large difference from the average value of F1, however for the F2 deviation is even less. but in the case of dataset 2 the difference for F1 is less compared to F2 difference.
For *Vowel u*, in the case of Dataset 1, the F1 and F2 difference is higher than difference of Dataset2
For *Vowel aa*, the F1 and F2 difference of Dataset 1is higher than difference of Dataset2 just as for vowel i.
For *Vowel ae*, the difference of the formants that we obtain in dataset 1 is lesser than for dataset 2. Thus we observe that the formant values of the centroid that we obtain by clustering depends on the dataset, the number of iterations the clustering is performed for. Also the difference that is being calculated is a function what is the average value of the formants tat we assume.

3 Appendix

Code
[mic1,fs]=audioread('cleanspeech.wav');

```
s1=size(mic1);
t = round(linspace(1,256,256));
N=256;
table = [];
z = 1;
M=s1(1);
K=fix(M/N);
preemph=[1 0.63];
for i=1:(2*K-1)
n =(1:N)+(N*(i-1)/2);
x1=mic1(n);
x=x1.*hamming(N);
x=filter(1,preemph,x);
X=fft(x);
[a,e]=lpc(x,10)
Xlpc=freqz(1,a);
rr=roots(a);
norm_freq=angle(rr);
freq_Hz=(norm_freq*fs)/(2*pi)

freq_Hz1 = abs(freq_Hz);
freq_Hz1 = sort(freq_Hz1);
pause
subplot(2,2,1);plot(20*log10(abs(X(1:N/2))))
subplot(2,2,2);plot(20*log10(abs(Xlpc)))
title(['Frame: ' num2str(i)]);
subplot(2,2,3);plot(x);
subplot(2,2,4);zplane(1,a)
pause
table(1,i) = i;
table(2,z) = freq_Hz1(1);
table(3,z) = freq_Hz1(3);
table(4,z) = freq_Hz1(5);
z = z+1;
end

i =1;

for i=1:z-1
temp = 0;
if table1(2,i) == 0
table1(2,i) = table1(4,i);
end
if table1(2,i)¿=table1(3,i)
temp = table1(2,i);
table1(2,i) = table1(3,i);
table1(3,i) = temp;
end
end

figure();
plot(table1(1,:),table1(2,:));
hold on
plot(table1(1,:),table1(3,:));
title('Frequencies of Formants 1 and 2 versus Frame Number');
xlabel('Frame Number'); ylabel('Frequency');
legend('Formant 1', 'Formant 2');
```