EE-414 Speech Processing Lab

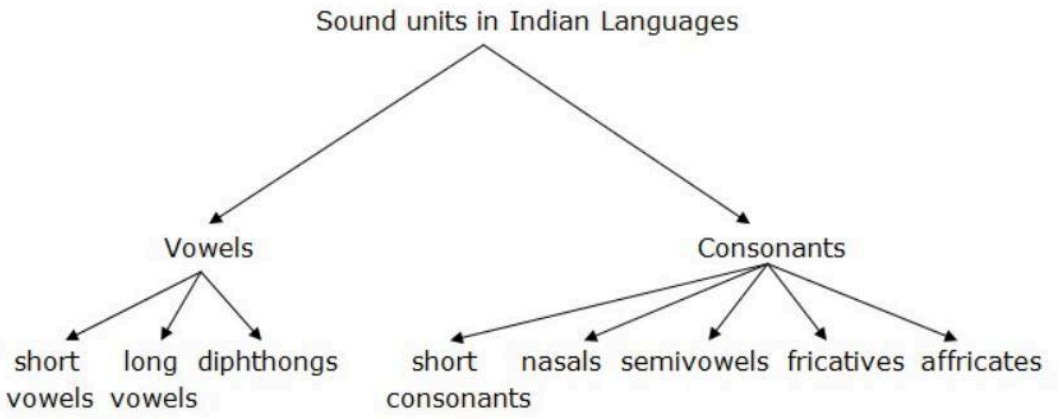
Lab-5-Phase-1 - 180020002

Aim:

- To study different sound units present in majority of Indian languages.
- To understand the production mechanism of each sound unit.
- To learn the time domain and frequency domain characteristics of different sound units.

Theory:

Classification of Sound units in Indian Languages:



Vowels & Consonants:

The sound units of most languages in India are broadly classified into two categories, namely, vowels and consonants. These two broad categories are mainly based on the shape of the vocal tract. In case of vowels, the vocal tract shape is wide open without any constriction along its length starting from the glottis till the lips and is excited by voiced excitation source. Alternatively, in case of consonants, their may be constriction in vocal tract shape some where along its length and is excited by either voiced, unvoiced and both types of excitation.

Short vowels, Long vowels and Diphthongs:

In most of the Indian languages, the vowel sound units are further classified into three categories as short vowels, long vowels and diphthongs. From the production process point of view there is no distinction between short and long vowels, except that the duration of production will be longer, typically nearly double that of short vowels. In case of diphthongs, as the name indicates, two vowel sounds are produced in succession without

any pause. The production process is such that the vocal tract shape is initially producing the first vowel and midway during the production of the first vowel it changes the shape to produce the other vowel .

Stop consonants:

Stop consonants form the major category of consonants in Indian languages. During the production of these consonants the vocal tract is completely closed at some point, somewhere along the length of the vocal tract and suddenly released. Hence the name stop consonants. The stop consonants are further classified into different cases based on two criteria, namely, place of articulation (POA) and manner of articulation(MOA). The POA gives the portion along the length of the vocal tract where it is completely closed. MOA gives the manner used for exciting the vocal tract synthesis, namely, voicing and aspiration.

In majority of the Indian languages we use voicing and aspiration as manner of articulation. Accordingly we have four possibilities, unvoiced unaspirated(UVUA), unvoiced aspirated(UVA), voiced unaspirated (VUA) and voiced aspirated(VA). The POA has typically five places of articulation in majority of the Indian languages. They are velar, palatal, alveolar, dental and bilabial. Among these the palatal POA stop consonants are categorized separately as affricates that will be described later. Thus we have four POA for the sounds of stop consonants. The stop consonants can also be grouped into four categories based on the POA. Accordingly in total we have about 4X4=16 stop consonants. Each of these stop consonants can be uniquely described in terms of MOA ans POA categories. For instances, the stop consonants |k| is characterized by UVUA MOA & velar POA. Further, the subset of stop consonants under the same category of MOA & POA will have same excitation characteristics and places, respectively. For instances, all stop consonants under UVUA MOA will have same MOA i.e. unvoicing & unaspiration.

MOA:

Vibration of vocal folds is a major excitation source during speech production. However, sound units, specifically, consonants may be produced using other types of unvoiced excitation like burst and frication. Accordingly the excitation can be either voiced and unvoiced. Further, aspiration is an important MOA in majority of Indian languages. The sound units produced using these MOA have different meanings.

POA:

For the production of stop consonants we will obstruct the vocal tract at different places along its length. These places are termed as POA. Stop consonants are therefore classified based on the POA also. As the different POA names given in Table indicates, the POA will be in velar, alveolar, dental & bilabial regions.

Velar Stop Consonants:

In this category the total constriction for the production of stop consonants occur at the velar region. In case of UVUA velar stop consonants |k|, there is no voicing and also aspiration. The only excitation is a burst of small duration, typically of about 5-20 msec. The two events that are present in the UVUA velar stop consonants is a closure region during which there is no speech activity and then the burst region where there will be sudden release of the closure.

Compare to UVUA velar stop consonants, UVA stop consonants will have three events, namely, closure & burst as in the case of UVUA & also aspiration event in addition. The closure & burst events will have similar class as in the UVUA case. The aspiration is due to the frication at the glottis region. The aspiration region can be identified as the noise like region after the burst region.

The VUA velar stop consonants will be same as UVUA, except that the unvoicing is replaced by a voiced excitation. Due to this there will be a low amplitude voicing bar in both the closure and burst regions.

The VA velar stop consonants is same as UVA velar stop consonants, except that unvoicing is replaced by a voicing process. Due to this there will be a voicing bar in the closure, burst & aspiration regions.

Alveolar Stop Consonants:

The total constriction of the vocal tract occurs at the alveolar ridge in this case. This category will also have four different stop consonants based on the MOA. The alveolar UVUA stop consonants production will be same as that of velar UVUA stop consonants except that the place of constriction is at the alveolar ridge. The change in the place of constriction will have effect mainly on the manifestation of burst region. The energy associated with the burst region & hence the prominence depends on the POA. The POA will in turn give knowledge how much burst oral cavity is present after release of is this cavity, then will be the energy associates with the burst & its prominence. Accordingly the burst region in case of alveolar stop consonants is relatively less prominent compared to the burst in case of velar stop consonants.

Dental Stop Consonants:

The POA will be dental region. The frontal cavity after the constriction will be further less and negligible compared to alveolar case. Hence the manifestation of burst region is very feeble in case of dental stop consonants. Apart from this there is no difference in the production compared to alveolar and velar stop consonants.

Bilabial stop consonants:

The POA is the two lips region & hence the name in the case of bilabial stop consonants. Since there is no frontal cavity after the POA, the burst region is not manifestated in case of bilabial stop consonants. However, the total closure & release of leads to the percception of bilabial stop consonants.

Procedure:

A. Short vowels, Long vowels and Diphthongs

a. Record the sounds of any one short vowel sound, long vowel sound and a diphthong (Also, record the two sounds present in the diphthong).

b. Plot the time domain waveform, magnitude spectrum and the spectrogram for each of the above sounds.

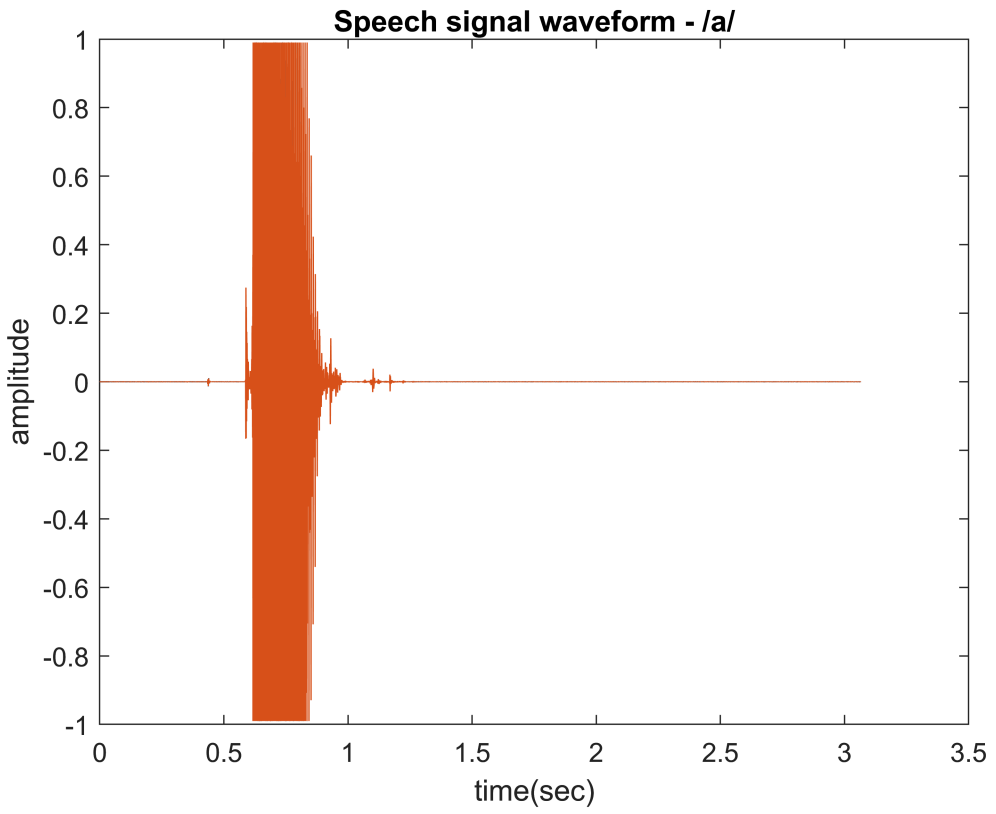
c. Inspect each of the above plots and write your observations comparing them.

PART : A

We can use audacity software to record the sounds and convert into sampling frequency of 16kHz and bit resolution as 16bits/sample. Then here we plot the waveforms.

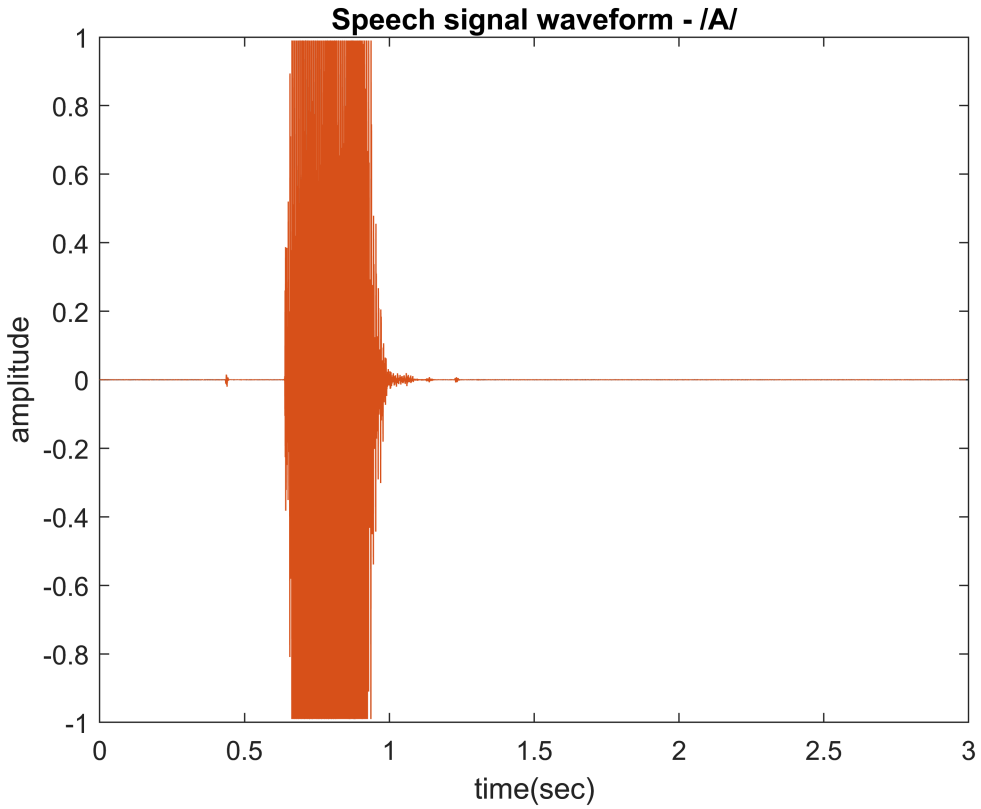
Time domain waveforms:

```
%Matlab program to load and plot time domain waveform stored in
% wav file format
%file name is l5_a.wav and full path is given for /a/ sound (short vowel)
[y,fs]=audioread('l5_a.wav');

%normalising the signal amplitudes to be in -1 to 1
y_a=y./(1.01*abs(max(y)));
%plotting waveform of the speech signal
t = 0 : 1 / fs : (length(y_a) - 1) / fs;
plot(t, y_a);
xlabel('time(sec)');
ylabel('amplitude');
title('Speech signal waveform - /a/');
```

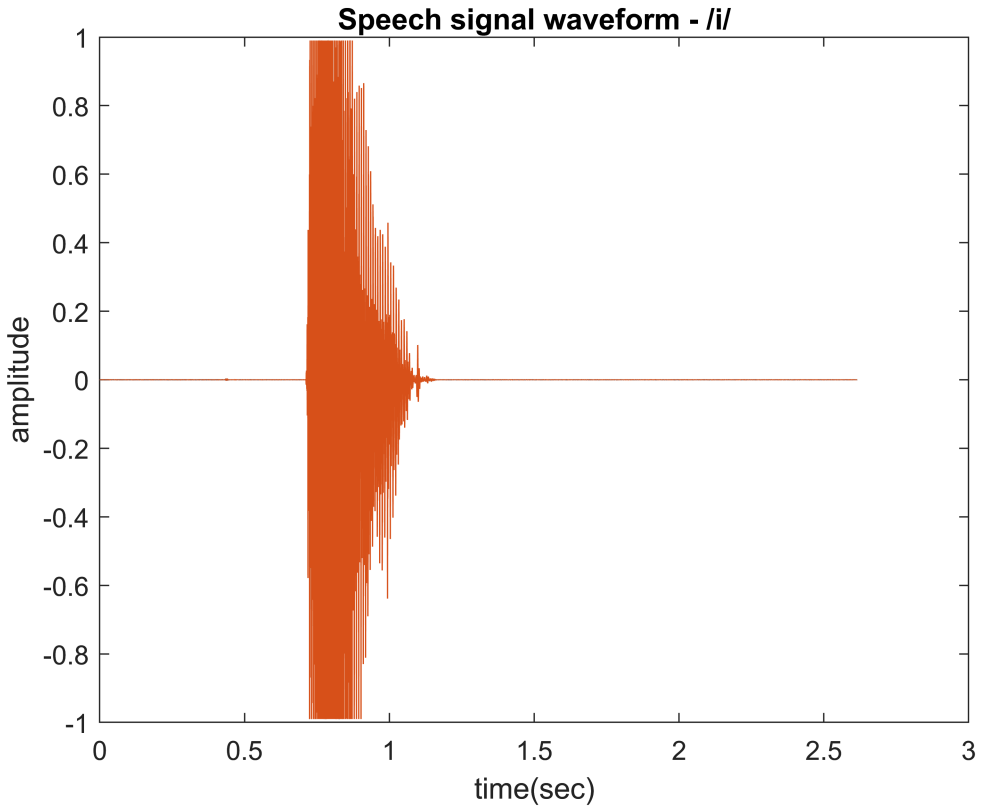



```
%file name is lab5_A.wav and full path is given for /A/ sound (long vowel)
[y,fs]=audioread('lab5_A.wav');

%normalising the signal amplitudes to be in -1 to 1
y_A=y./(1.01*abs(max(y)));
%plotting waveform of the speech signal
```

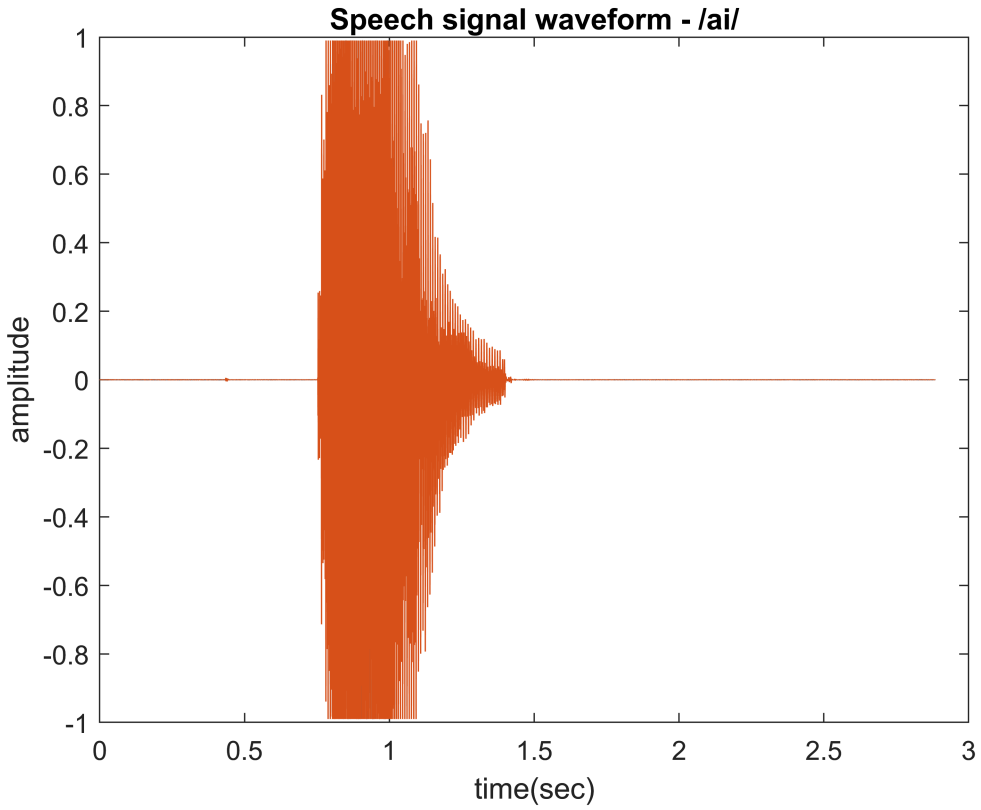
```
t = 0 : 1 / fs : (length(y_A) - 1) / fs;
plot(t, y_A);
xlabel('time(sec)');
ylabel('amplitude');
title('Speech signal waveform - /A/');
```




```
%file name is l5_i.wav and full path is given for /i/ sound (short vowel)
[y,fs]=audioread('l5_i.wav');

%normalising the signal amplitudes to be in -1 to 1
y_i=y./(1.01*abs(max(y)));
%plotting waveform of the speech signal
t = 0 : 1 / fs : (length(y_i) - 1) / fs;
plot(t, y_i);
xlabel('time(sec)');
ylabel('amplitude');
title('Speech signal waveform - /i/');
```




```
%file name is l5_ai.wav and full path is given for /ai/ sound (diphthong)
[y,fs]=audioread('l5_ai.wav');

%normalising the signal amplitudes to be in -1 to 1
y_ai=y./(1.01*abs(max(y)));
%plotting waveform of the speech signal
t = 0 : 1 / fs : (length(y_ai) - 1) / fs;
plot(t, y_ai);
xlabel('time(sec)');
ylabel('amplitude');
title('Speech signal waveform - /ai/');
```



Magnitude spectrum:

We will use the wavesurfer for analysing the plots. We can note down the 25ms duration of the segment at the centre of the sound. The time-stamps for each sound is obtained from wavesurfer.

```
%/a/
y_a = y(ceil(0.6655*fs) : floor(0.6905*fs));
%/A/
y_A = y(ceil(0.7694*fs) : floor(0.7944*fs));
%/i/
y_i = y(ceil(0.8426*fs) : floor(0.8676*fs));
%/ai/
y_ai = y(ceil(1.056*fs) : floor(1.081.*fs));
```

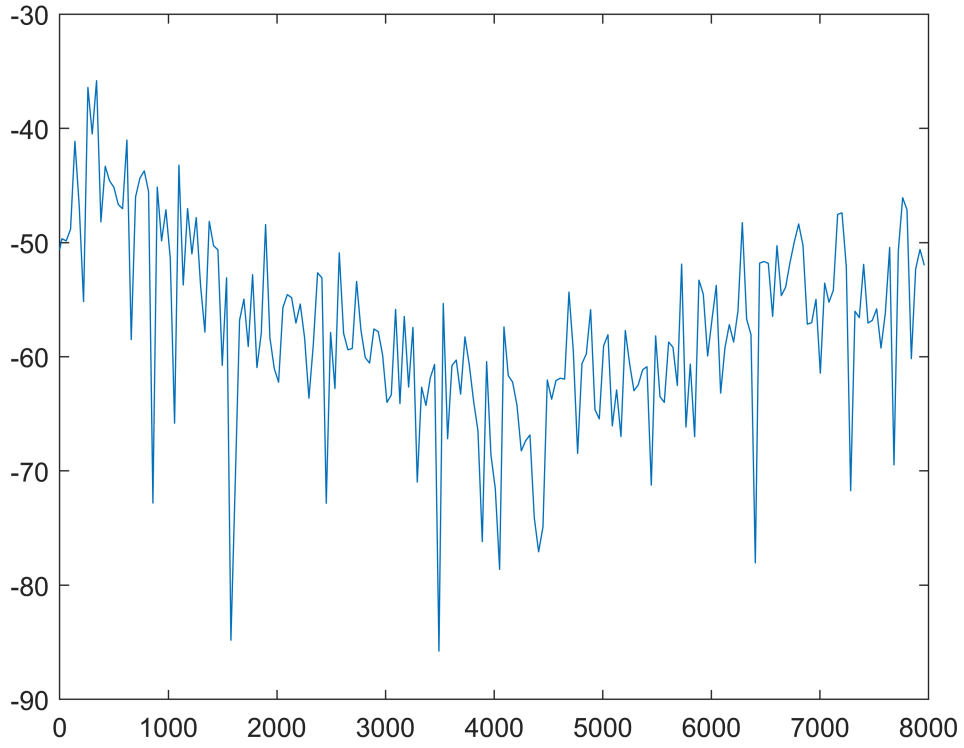
Now we plot the magnitude spectrum plots of the speech signal. We use the same method as specified to compute the N-point DFT.

```
Y_a = fftshift(fft(y_a));
Y_A = fftshift(fft(y_A));
Y_i = fftshift(fft(y_i));
Y_ai = fftshift(fft(y_ai));
```

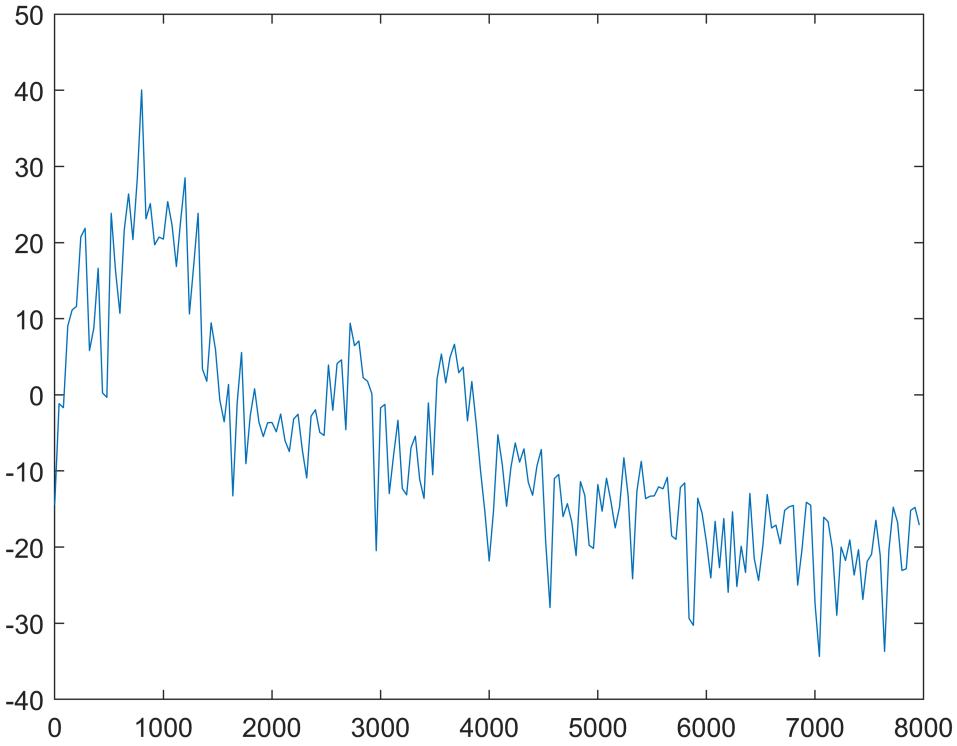
We have obtained the N-point DFTs of all the sounds. Now we plot the frequency spectrum for positive frequencies.

```
F_a = -fs/2 : fs/length(Y_a) : fs/2 - fs/length(Y_a);
```

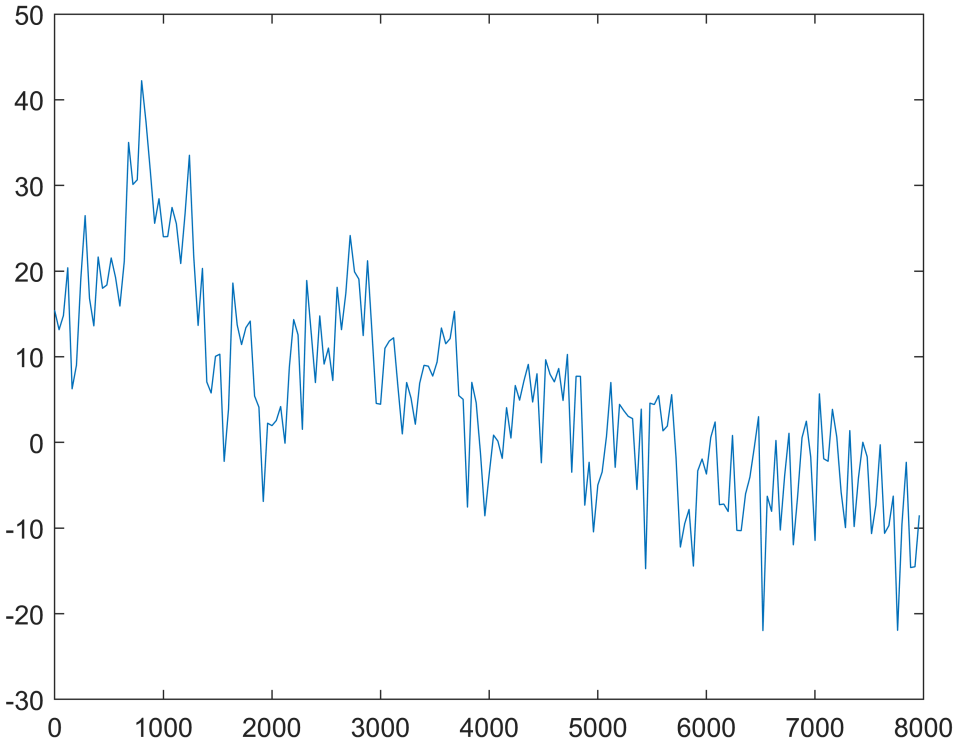
```
F_A = -fs/2 : fs/length(Y_A) : fs/2 - fs/length(Y_A);
F_i = -fs/2 : fs/length(Y_i) : fs/2 - fs/length(Y_i);
F_ai = -fs/2 : fs/length(Y_ai) : fs/2 - fs/length(Y_ai);

% Plots
% /a/
plot(F_a, 20*log10(abs(Y_a)));
xlim([0, fs/2]);
```



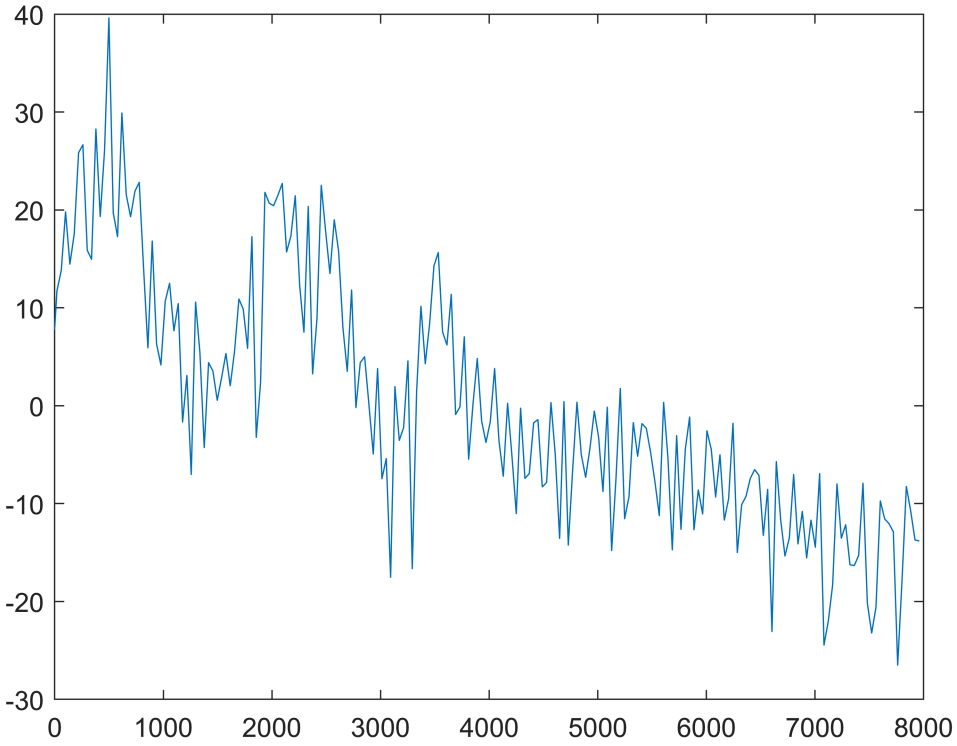

```
% /A/
plot(F_A, 20*log10(abs(Y_A)));
xlim([0, fs/2]);
```




```
% /i/
plot(F_i, 20*log10(abs(Y_i)));
xlim([0, fs/2]);
```

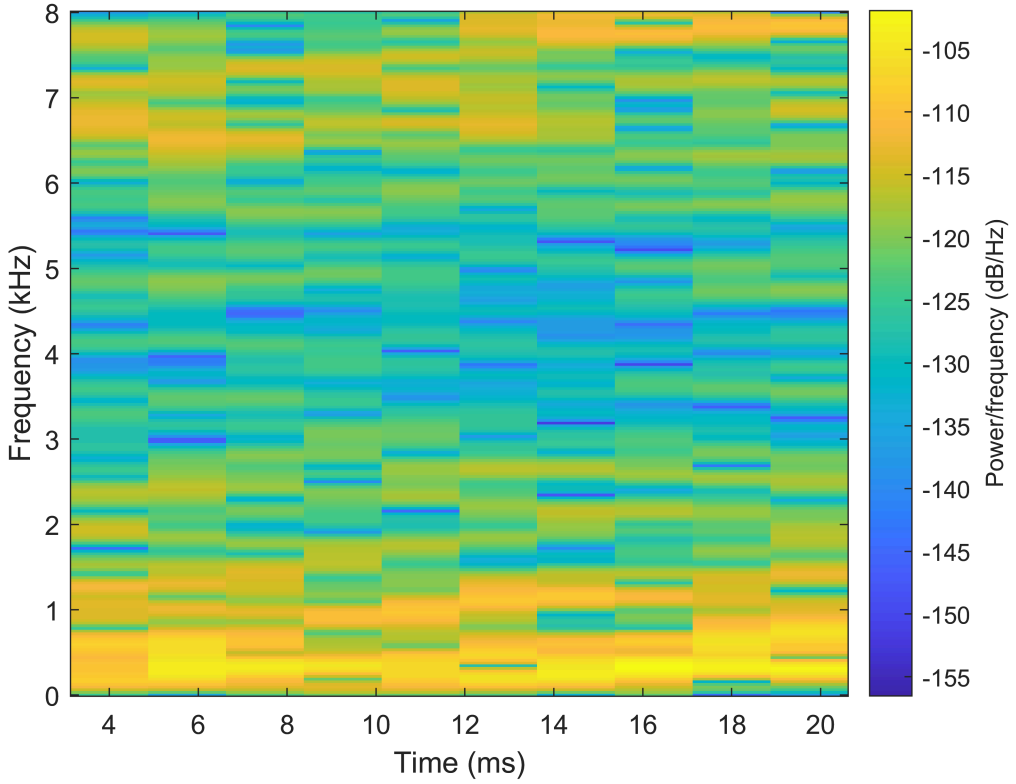



```
% /ai/
plot(F_ai, 20*log10(abs(Y_ai)));
xlim([0, fs/2]);
```

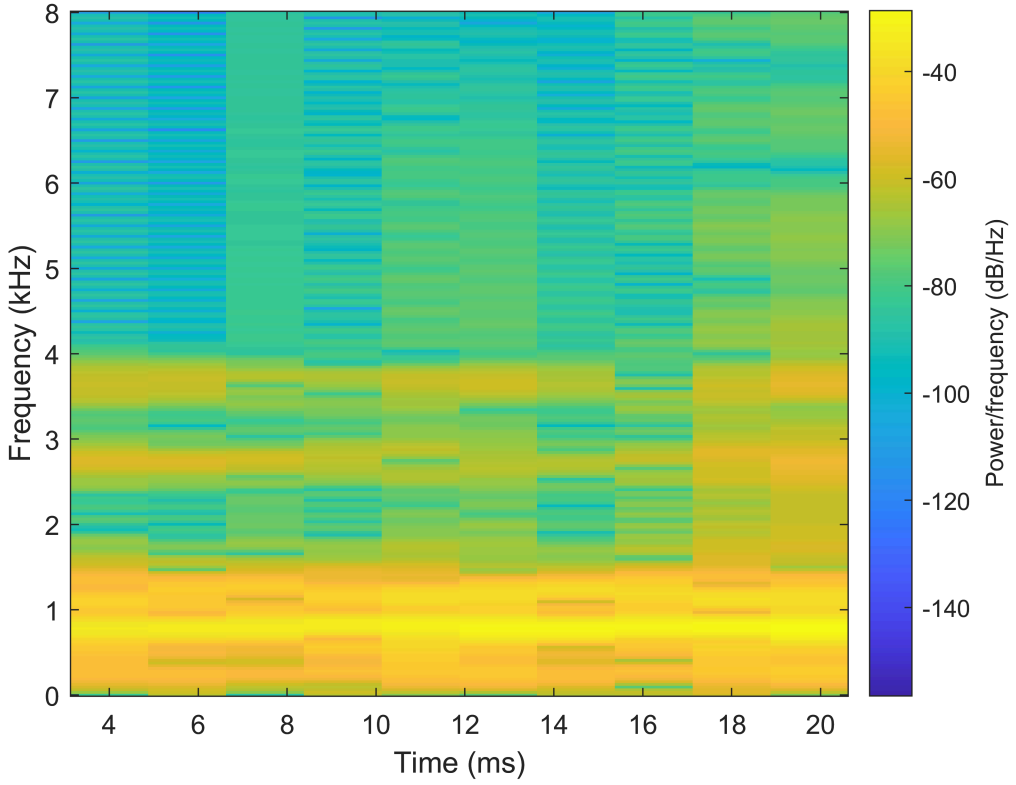


Spectrogram:

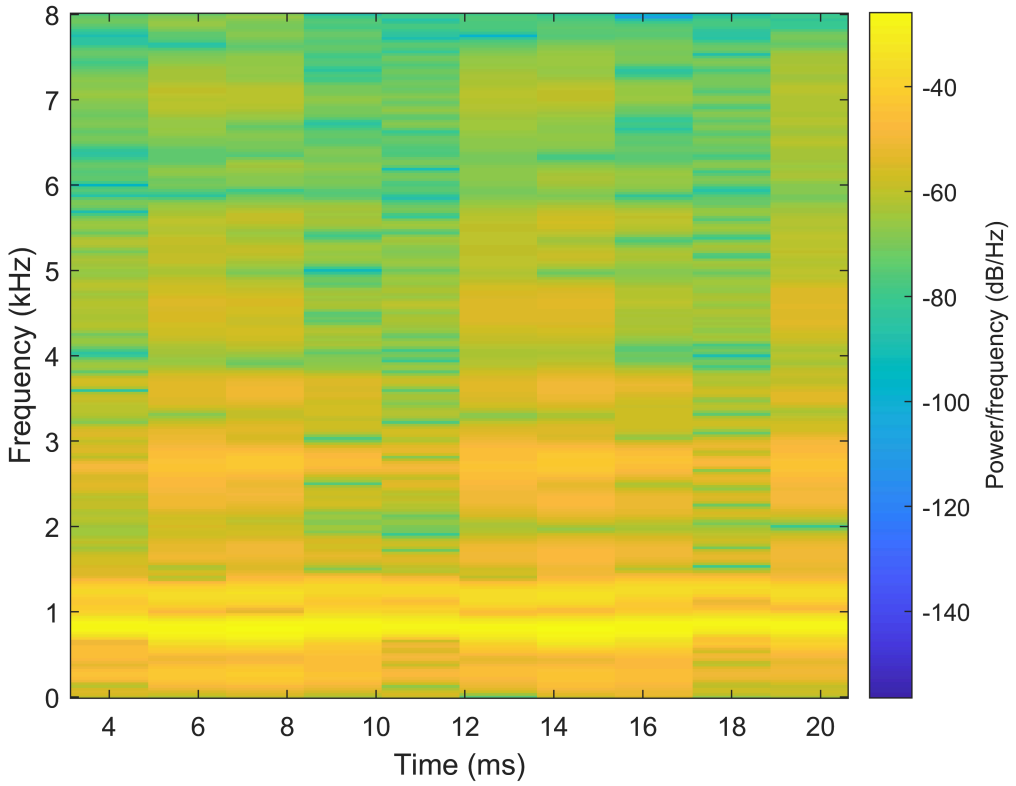
```
spectrogram(y_a,  128, (100), 512, fs,'yaxis');
```



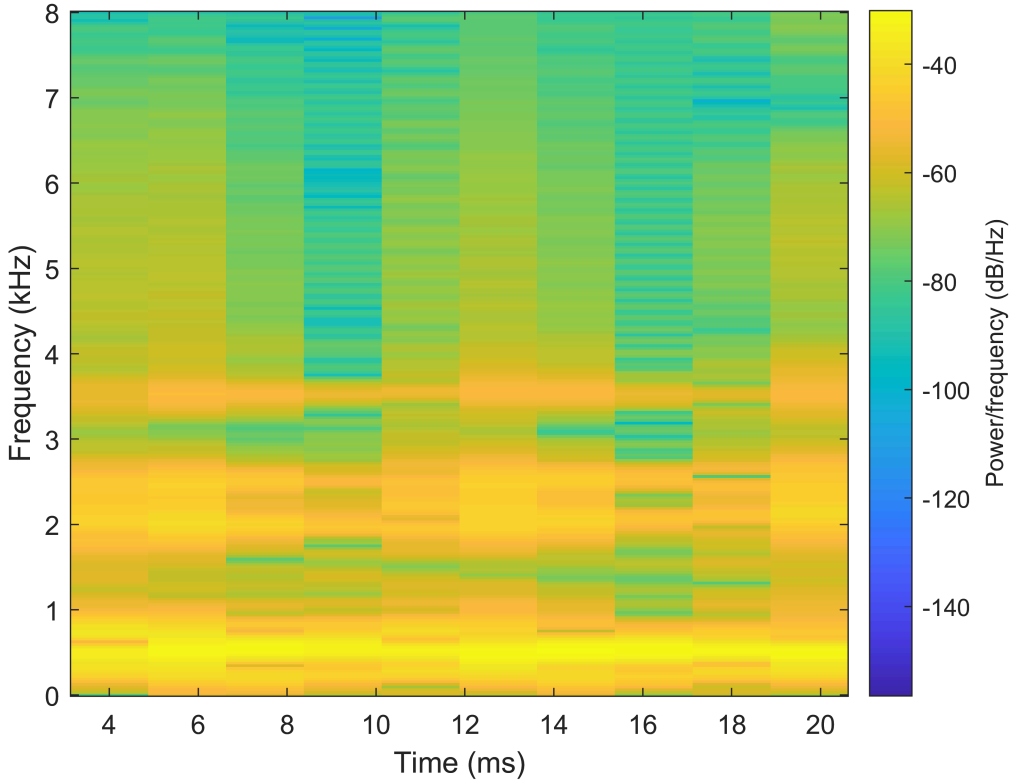

```
spectrogram(y_A,  128, (100), 512, fs,'yaxis');
```



```
spectrogram(y_i,  128, (100), 512, fs,'yaxis');
```




```
spectrogram(y_ai,  128, (100), 512, fs,'yaxis');
```



Observations:

From the production process point of view there is no distinction between short and long vowels, except that the duration of production will be longer in long vowels than short vowels. Here, we can observe from the time domain plots that /A/ has longer time duration than /a/. But the magnitude spectrum and spectrogram is same for /a/ & /A/. The periodic pattern and energy levels are same in /a/ and /A/.

The production process in diphthongs is such that the vocal tract shape is initially producing the first vowel and midway during the production of the first vowel it changes the shape to produce the other vowel.

In case of diphthongs /ai/, the initial portion resembles with /a/ and the later portion with /i/. Transition of vocal tract shape from |a| to |i| can be clearly seen in the spectrogram. Periodicity and energy level changes during the transition from /a/ to /i/.

B. Stop Consonants

POA	MOA			
	UVUA	UVA	VUA	VA
Velar	k	k^h	g	g^h
Aveolar	T	T^h	D	D^h
Dental	t	t^h	d	d^h
Bilabial	p	p^h	b	b^h

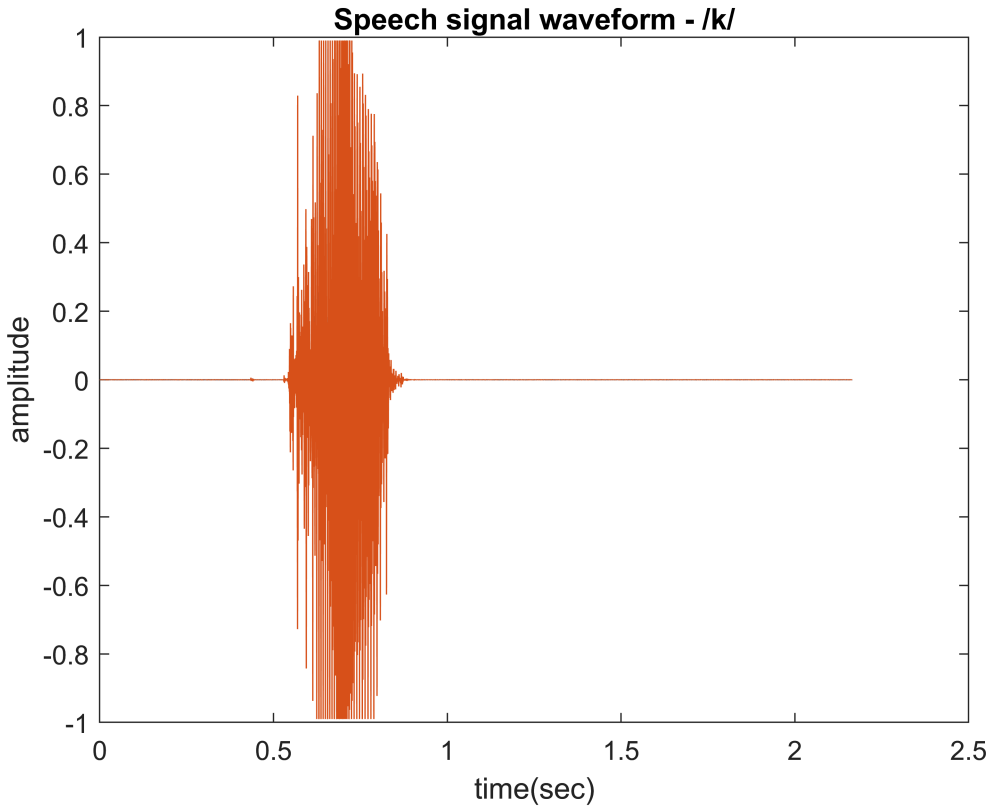
a. Pick up any one of the POA(Position of Articulation) types and record the sounds present in the respective row for all the MOA(Manner of Articulation) types.
b. Plot the time domain waveform, the magnitude spectrum and the spectrogram for each of the above sounds.
c. Inspect the above plots and describe the various sub phonetic events that take place, their relative duration and how they vary across different kinds of MOA.

PART : B

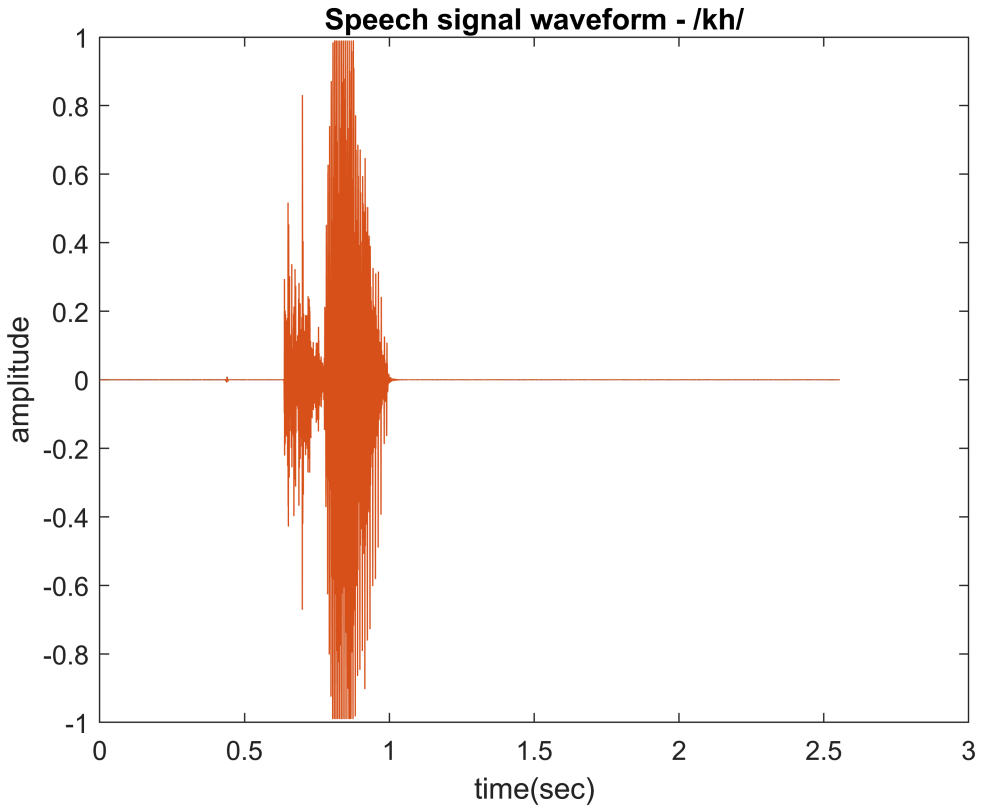
We can use audacity software to record the sounds and convert into sampling frequency of 16kHz and bit resolution as 16bits/sample. Then here we plot the waveforms.

Time domain waveforms:

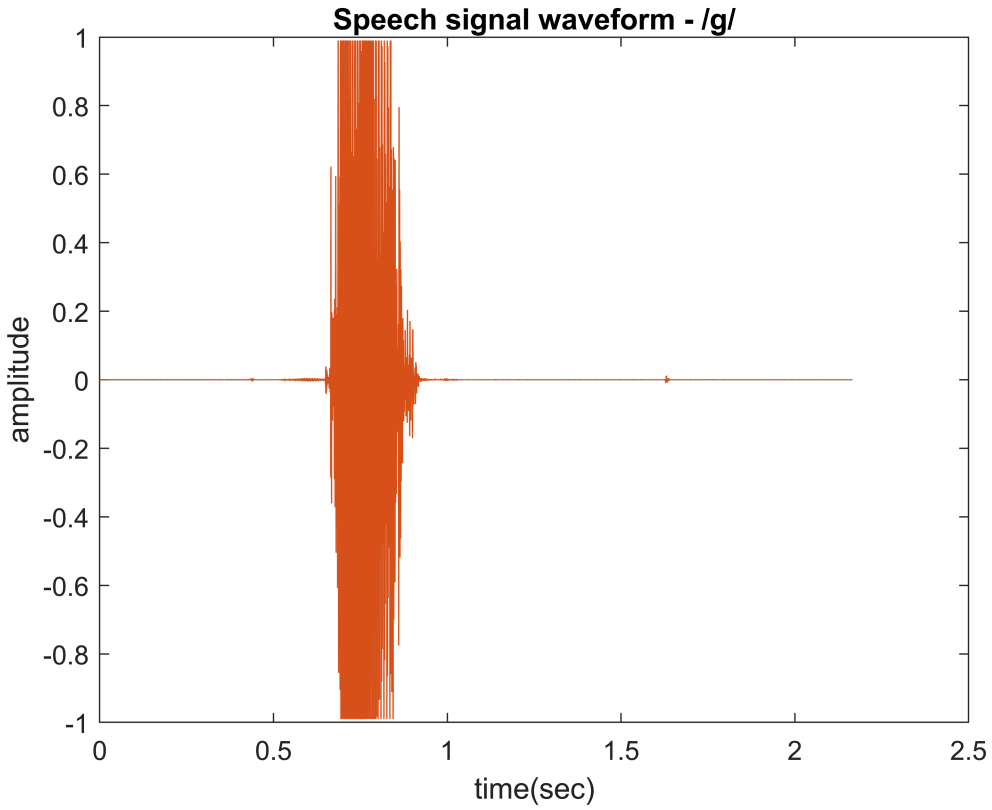
```
%Matlab program to load and plot time domain waveform stored in
% wav file format
%file name is l5_k.wav and full path is given for /k/ sound (Velar-UVUA)
[y,fs]=audioread('l5_k.wav');

%normalising the signal amplitudes to be in -1 to 1
y_k=y./(1.01*abs(max(y)));
%plotting waveform of the speech signal
t = 0 : 1 / fs : (length(y_k) - 1) / fs;
plot(t, y_k);
xlabel('time(sec)');
ylabel('amplitude');
title('Speech signal waveform - /k/');
```



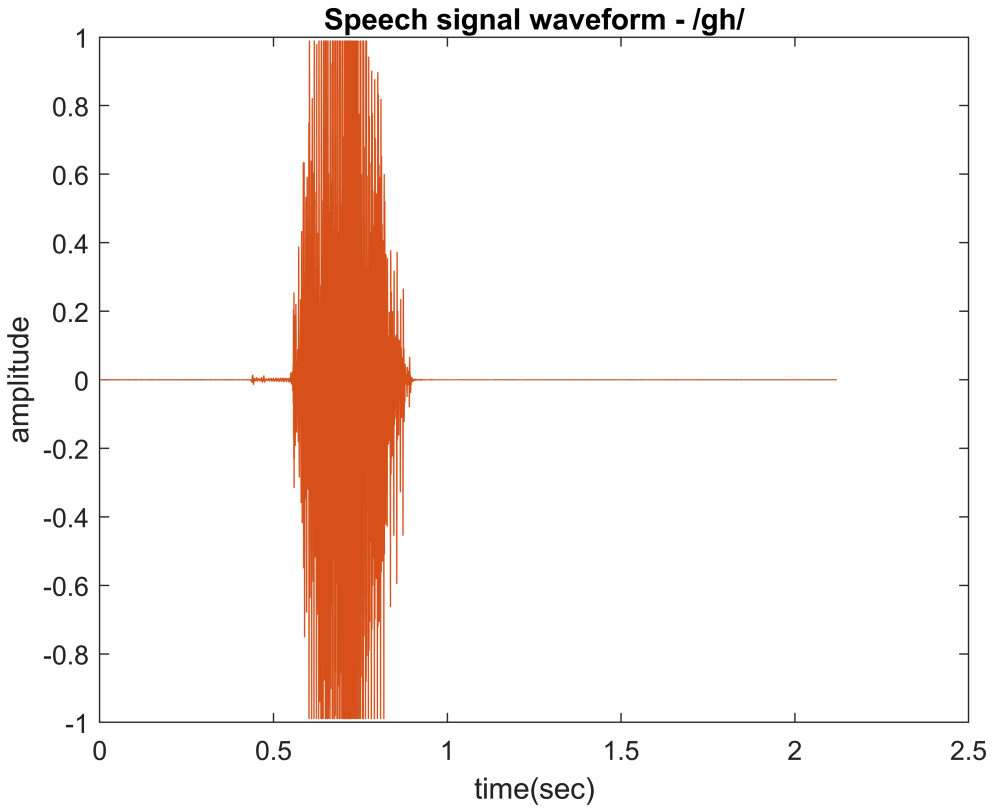

```
%file name is l5_kh.wav and full path is given for /kh/ sound (Velar-UVA)
[y,fs]=audioread('l5_kh.wav');

%normalising the signal amplitudes to be in -1 to 1
y_kh=y./(1.01*abs(max(y)));
%plotting waveform of the speech signal
t = 0 : 1 / fs : (length(y_kh) - 1) / fs;
plot(t, y_kh);
xlabel('time(sec)');
ylabel('amplitude');
title('Speech signal waveform - /kh/');
```




```
%file name is l5_g.wav and full path is given for /g/ sound (Velar-VUA)
[y,fs]=audioread('l5_g.wav');

%normalising the signal amplitudes to be in -1 to 1
y_g=y./(1.01*abs(max(y)));
%plotting waveform of the speech signal
t = 0 : 1 / fs : (length(y_g) - 1) / fs;
plot(t, y_g);
xlabel('time(sec)');
ylabel('amplitude');
title('Speech signal waveform - /g/');
```




```
%file name is l5_gh.wav and full path is given for /gh/ sound (Velar-VA)
[y,fs]=audioread('l5_gh.wav');

%normalising the signal amplitudes to be in -1 to 1
y_gh=y./(1.01*abs(max(y)));
%plotting waveform of the speech signal
t = 0 : 1 / fs : (length(y_gh) - 1) / fs;
plot(t, y_gh);
xlabel('time(sec)');
ylabel('amplitude');
title('Speech signal waveform - /gh/');
```



Magnitude spectrum:

We will use the wavesurfer for analysing the plots. We can note down the 25ms duration of the segment at the centre of the sound. The time-stamps for each sound is obtained from wavesurfer.

```
%/k/
y_k = y(ceil(0.6648*fs) : floor(0.6898*fs));
%/kh/
y_kh = y(ceil(0.8716*fs) : floor(0.8966*fs));
%/g/
y_g = y(ceil(0.739*fs) : floor(0.764*fs));
%/gh/
y_gh = y(ceil(0.6573*fs) : floor(0.6823.*fs));
```

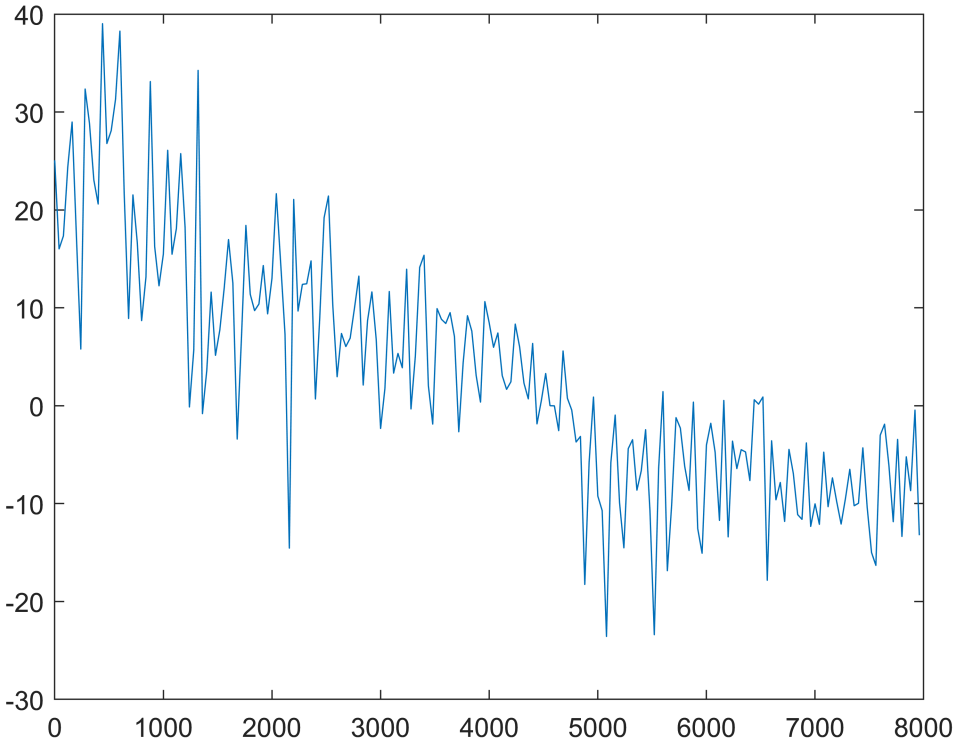
Now we plot the magnitude spectrum plots of the speech signal. We use the same method as specified to compute the N-point DFT.

```
Y_k = fftshift(fft(y_k));
Y_kh = fftshift(fft(y_kh));
Y_g = fftshift(fft(y_g));
Y_gh = fftshift(fft(y_gh));
```

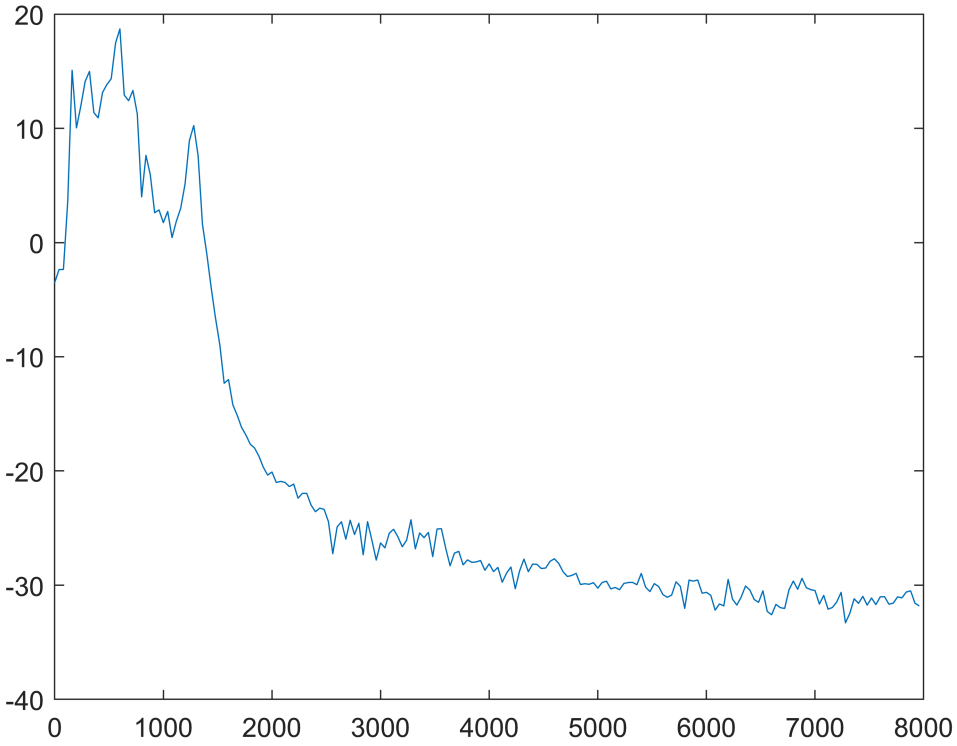
We have obtained the N-point DFTs of all the sounds. Now we plot the frequency spectrum for positive frequencies.

```
F_k = -fs/2 : fs/length(Y_k) : fs/2 - fs/length(Y_k);
```

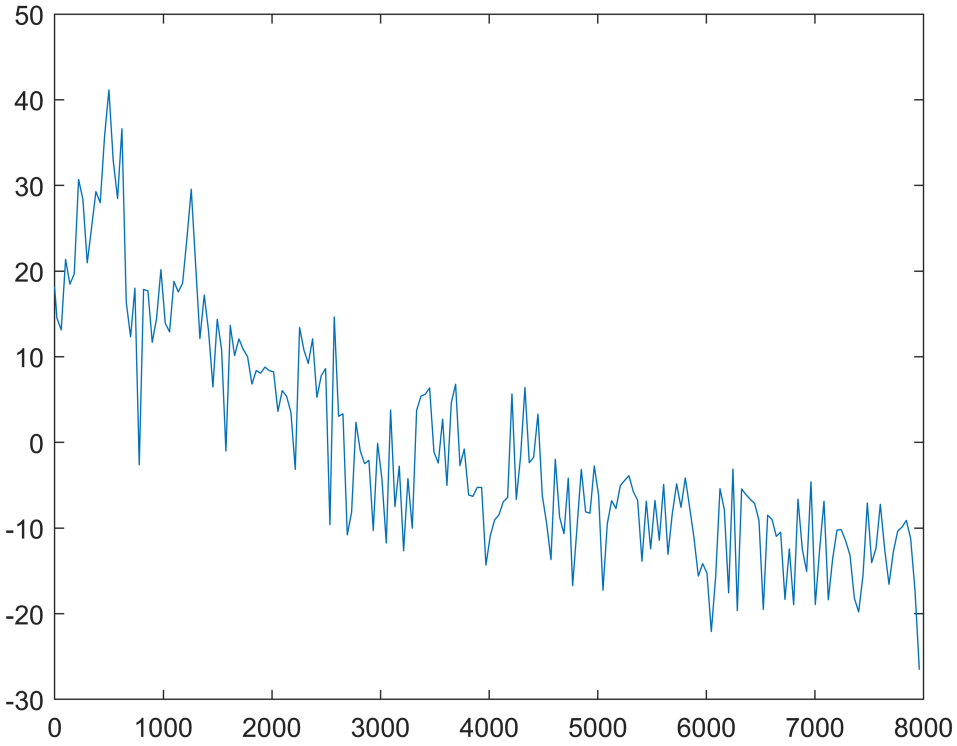
```
F_kh = -fs/2 : fs/length(Y_kh) : fs/2 - fs/length(Y_kh);
F_g = -fs/2 : fs/length(Y_g) : fs/2 - fs/length(Y_g);
F_gh = -fs/2 : fs/length(Y_gh) : fs/2 - fs/length(Y_gh);

% Plots
% /k/
plot(F_k, 20*log10(abs(Y_k)));
xlim([0, fs/2]);
```



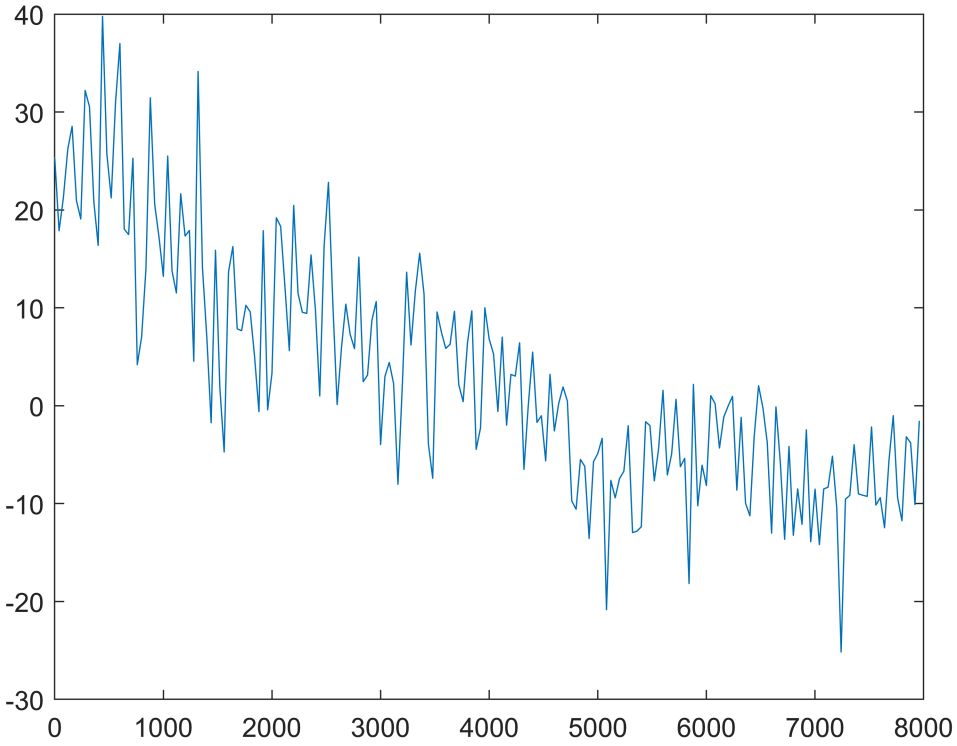

```
% /kh/
plot(F_kh, 20*log10(abs(Y_kh)));
xlim([0, fs/2]);
```




```
% /g/
plot(F_g, 20*log10(abs(Y_g)));
xlim([0, fs/2]);
```

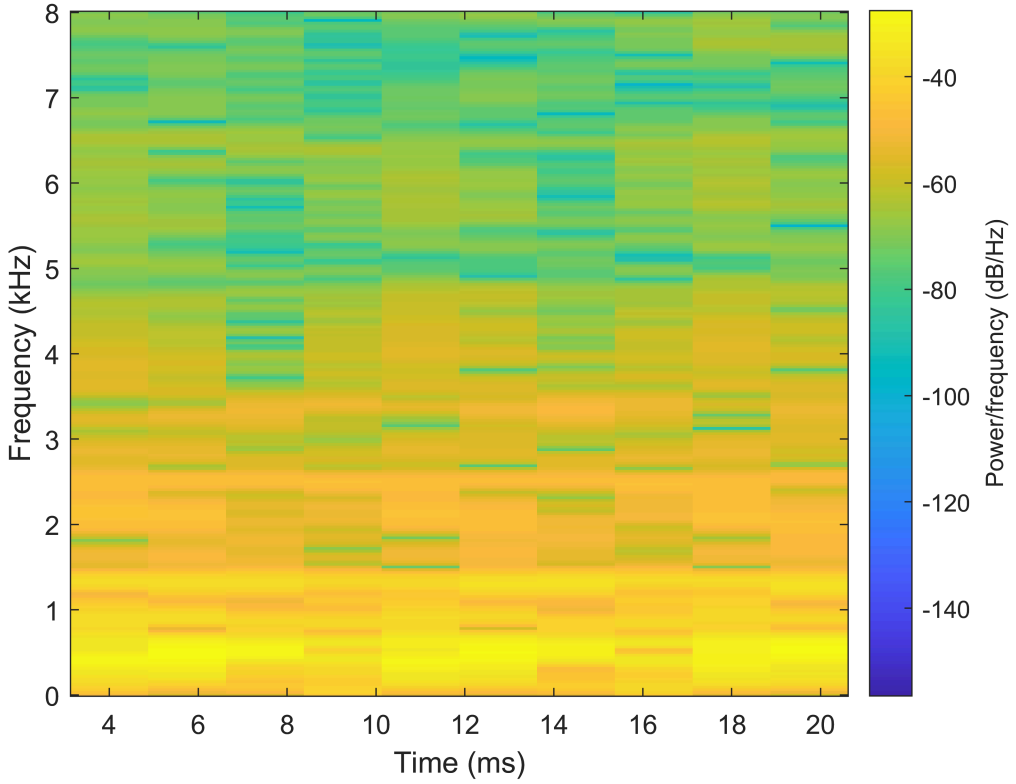



```
% /gh/
plot(F_gh, 20*log10(abs(Y_gh)));
xlim([0, fs/2]);
```

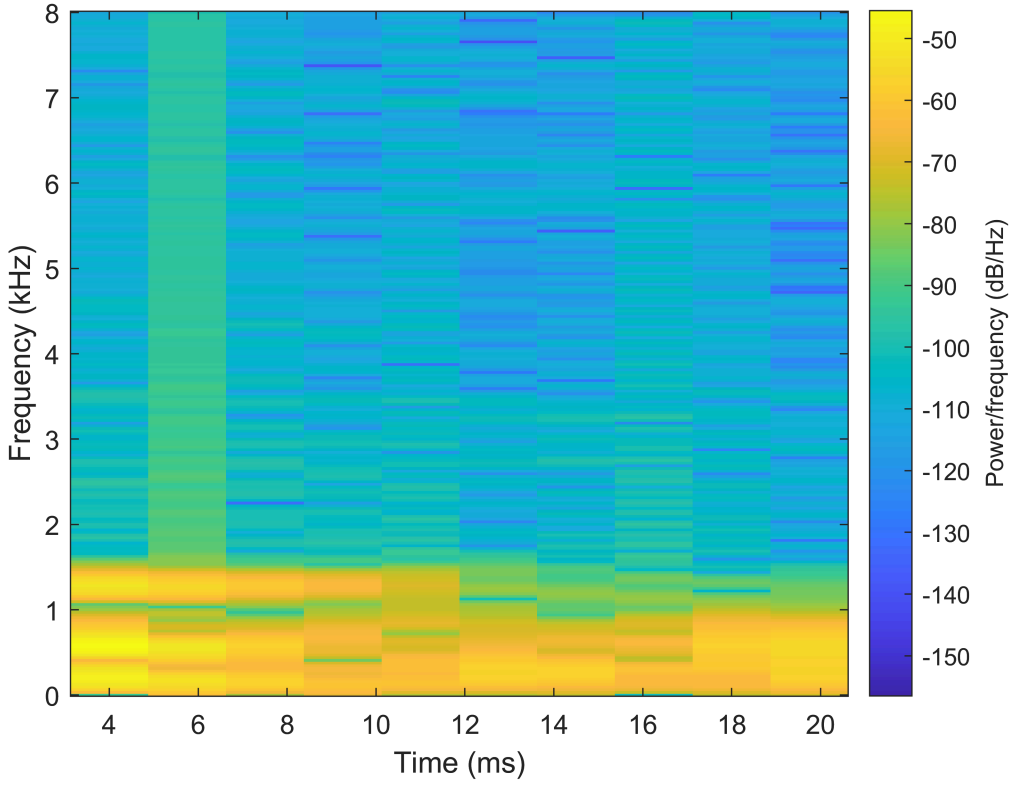


Spectrogram:

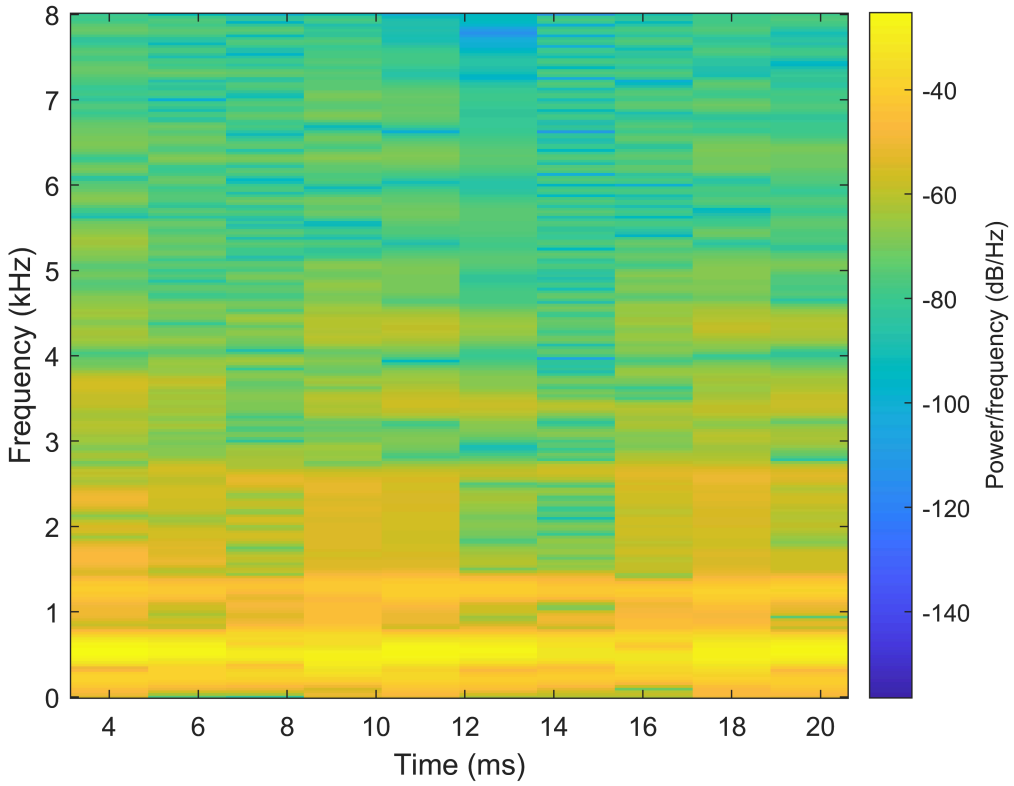
```
spectrogram(y_k,  128, (100), 512, fs,'yaxis');
```



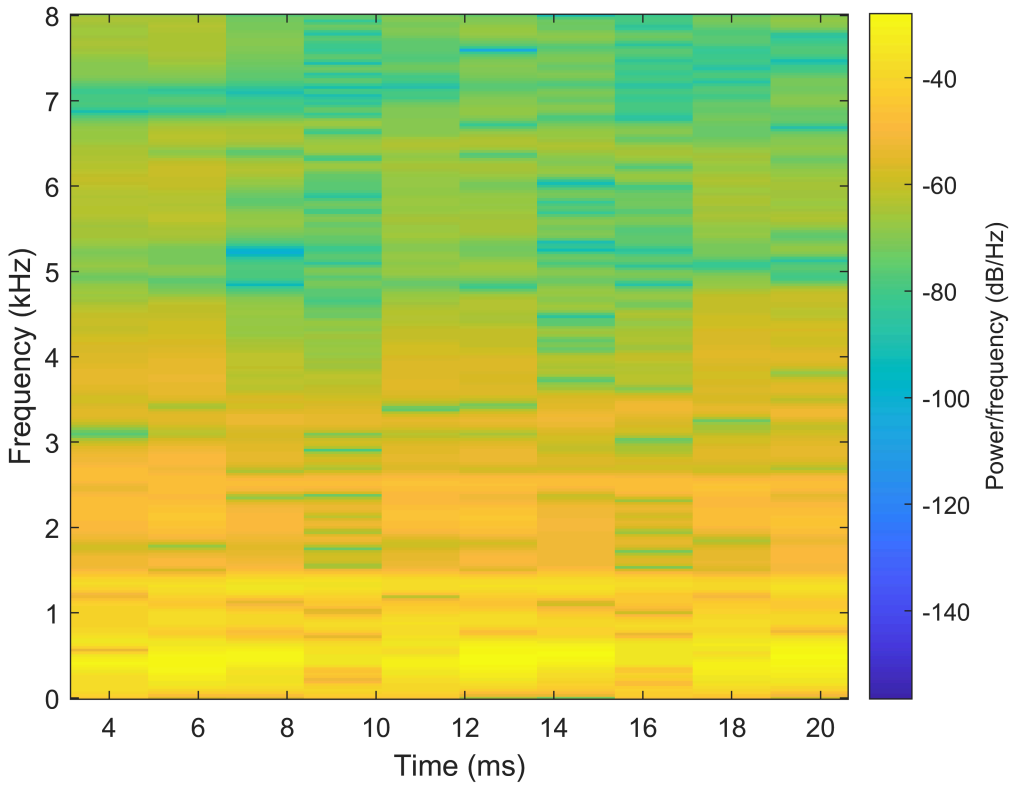

```
spectrogram(y_kh,  128, (100), 512, fs,'yaxis');
```



```
spectrogram(y_g,  128, (100), 512, fs,'yaxis');
```




```
spectrogram(y_gh,  128, (100), 512, fs,'yaxis');
```



Observations:

We didn't observe any periodicity in the /k/ and /kh/ as these consonants are unvoiced. Noise like waveform of the consonant /kh/ indicates the /kh/ is produced with aspiration whereas /k/ is unaspirated. The energy levels are high for unvoiced sounds here.

In /k/, the closure region is a silence region, difficult to distinguish from the non-speech region followed by a burst region.

In /kh/, the aspiration region can be identified as the noise like region after the burst region.

Also, /g/ has periodicity as it is voiced and unaspirated. But /gh/ doesnt have periodicity as it is produced with aspiration. We also observe noise like behaviour in the waveform. Here, the energy levels were low comparitively for voiced sounds.

In /g/, the only difference with reference to UVUA is the presence of low amplitude nearly periodic signal in the time domain. The low level voicing bar corresponds to the low frequency spectral energy in the frequency domain. The spectrogram shows two different spectral information, one representing the closure region & the other representing the burst region.

Now in /gh/, there is a voicing bar in the closure, burst & aspiration regions.