Audio Vowel Detection

# Title

Detect the vowels in speech signal along with harmonics and fundamental frequency.

# Introduction

The focus of this project is to detect the vowels from a high-quality audio file in MATLAB. The .wav file is loaded in MATLAB and its analysis is done in time and frequency domain. Then fundamental frequency and harmonics are traced out from the frequency domain spectrum of audio file. For vowel detection speech signal is analyzed in passband filter of frequency 0 to 4kHz. Half wave rectification is applied along with the low pass filter with a cut off frequency at 28Hz. Down sampling of audio file is done at 80 samples/sec with the normalization of the signal. Audio signal is discretized using DFT over hamming window. As vowels are the words contains high energy spectrum, so the next step is to unsampled energy components at the rate of 8000 samples/sec. For the enhancement of vowels onset point slopes are calculated. Again, rectification is applied after smoothing the signal. Then desirable peaks can be found. In this project, we have used vop detection method using energy spectrum.

# Mathematics involved

The equation for finding energy distribution of bands:

A picture containing text, watch

Description automatically generated

Where,



From this equation we get energy distribution, after this it is required to enhance the energy to smooth the signal and the output of VOP evidence is plotted.

# Implementation

**Audio files in MATLAB:**

Graphical user interface, application

Description automatically generated

Load the audio files in MATLAB using the audioread() function. Sample length of the audio is calculated. To obtain the plot of audio file in time domain, where t represents the time elapsed.

Graphical user interface, text, application

Description automatically generated

Sampling time is calculated, and length of each sample is also calculated. Figure 1 shows the time domain signal of audio file “Television.wav”. The speech signal is plotted in figure 1.

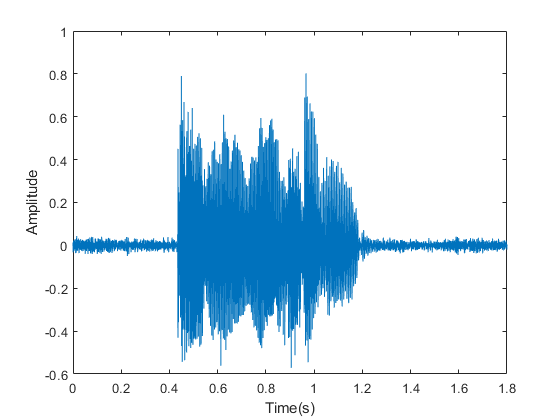
.

Figure 1 shows the time domain signal of audio file.

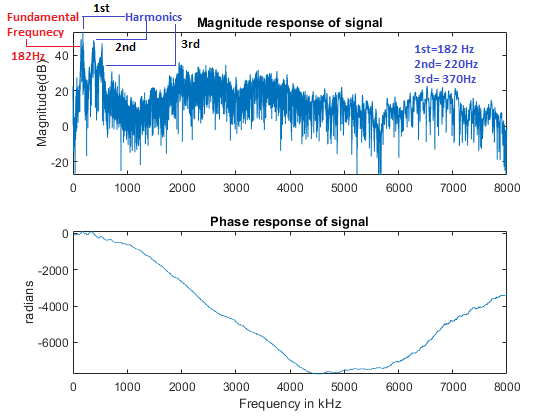


Figure shows the plot of audio file in frequency domain.

Then frequency domain plot is observed using fft(). Fast Fourier transform of audio signal is performed. Magnitude and phase response of audio file is generated. Figure 2 shows the audio signal is frequency domain. In frequency domain of the entire spectrum fundamental frequency is highlighted. First, second and third harmonics are labeled here.

**Frequency and harmonics analysis:**

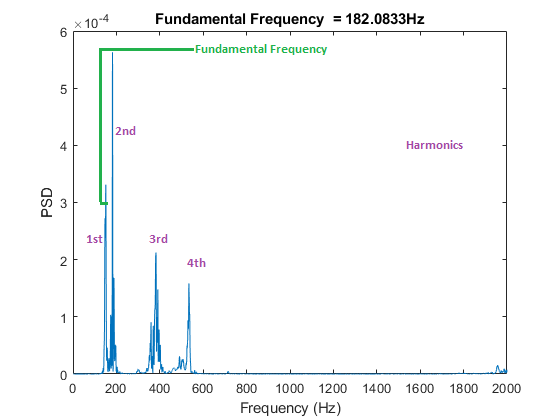


Figure shows the fundamental frequency and harmonics.

Figure 3 shows the fundamental frequency component of audio signal. It is calculated around 182Hz. The fundamental frequency is also the frequency of first harmonic. Harmonics in the signal are the integral multiple of fundamental frequency. In figure 3 harmonics can be seen at 182, 200, 400, 500 Hz. If we check back figure 2, here we can see there are peaks at the start of the spectrum. These frequencies are plotted separately here in figure 3. Harmonics may contain high energy spectrum. And vowels spoken also contains high energy than consonants. Also there harmonics matches the points of vowels in figure 4.

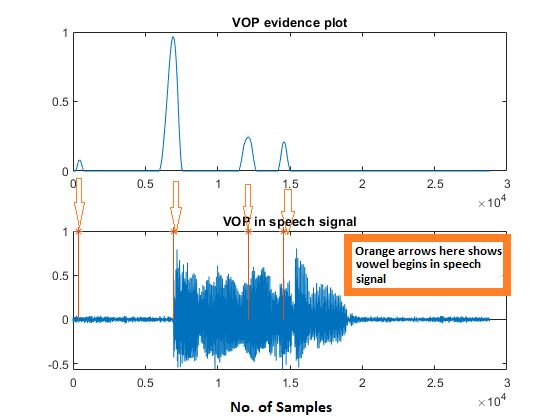


Figure shows the VOP energy spectrum detecting vowels.

In figure 4, Vop evidence plot shows peaks when the vowel is spoken. These plots are taken using the energy modulation spectrum. When vowels are spoken high energy is released. When high energy spectrum is detected, we get orange line. In second subplot() vowels are plotted over speech signal. Here orange line shows the VOP.

The main observation in figure 4 can be seen that we there are four number of vowels showing. The original file which have been loaded in MATLAB was television.wav. This word contain five vowels and we are detecting four vowels because vowel onset point indicate the starting point of vowel.

# Drawbacks

Noise can affect the results. If larger amplitude noise is added to the signal, it may not produce correct results.

In MATLAB, large audio files may cause delay in the results and ultimately shifting of VOP.

# Appendices

clc;

clear all;

close all;

%% audioread

[y,fs]=audioread('What\_are\_you\_doing\_here.wav'); %read audio file

%[y,fs]=audioread('na.wav');

N = length(y);

q=18;

%% filtfilting

p=4000/q;

x(1,:)=bandpass(y(:,1),[15 p],fs);

for i=2:q

x(i,:)=bandpass(y(:,1),[(i-1)\*p i\*p],fs);

end

%% half wave rectifier

for i=1:q

for j=1:N

if x(i,j)<0

a(i,j)=0;

else

a(i,j)=x(i,j);

end

end

end

%% Lowpass filtfilt

for i=1:q

lp(i,:)=lowpass(a(i,:),28,fs,'Steepness',0.95);

% lp(i,:)=smooth(lp(i,:),50\*fs/1000);

end

%% normalise

for i=1:q

dw(i,:)=lp(i,:)./max(lp(i,:));

end

%% fft window

ham=hamming(20);

for i=1:q

temp=dw(i,:);

temp=buffer(temp,20,19);

temp=temp.\*(ham\*ones(1,length(temp)));

temp=fft(temp,40);

temp = sum(abs(temp(4:16,:)));

temp1(i,:)=temp;

end

temp1=sum(temp1);

temp1=resample(temp1,80,fs);

temp1=resample(temp1,fs,80);

temp1=temp1/max(temp1);

%% enhancement

temp1=filtfilt(hamming(1600),1,temp1);

y1=diff(temp1);

y2=buffer(y1,160,159);

y3=sum(y2);

y4=y3;

y4(y3<0)=0;

Fogd=diff(gausswin(800));

y5=filter(Fogd,1,y4);

%% Enhancing the VOP and removing close peaks and ploting it with speech signal

y5=y5./max(y5);

y5(y5<0)=0;

y5(y5<0.1)=0;

y5=smooth(y5,320);

[pks, vop]=findpeaks(y5);

figure(1)

subplot(211)

plot(y5)

title('VOP evidence plot');

subplot(212)

plot(y-mean(y))

hold on

stem(vop,ones(size(pks)),'\*')

title('VOP in speech signal')

%% Time domain plot

sound(y,fs);

N = size(y,1);

len=length(y)/fs; %We get time of audio in seconds

sam\_len = floor(((len\*1000)- 10)/(10));

% max1= zeros(10);

getans=zeros(sam\_len,1);

figure(2)

t = (0:N-1)/fs;

plot(t, y)

xlabel('Time(s)')

ylabel('Amplitude')

%% Frequnecy domain plot

Y = fft(y,N);

F = ((0:1/N:1-1/N)\*fs).';

magnitudeY = abs(Y); % Magnitude of the FFT

phaseY = unwrap(angle(Y)); % Phase of the FFT

dB\_mag=mag2db(magnitudeY);

figure(3)

subplot(2,1,1);

plot(F(1:end/2),dB\_mag(1:end/2));

title('Magnitude response of signal');

ylabel('Magnitude(dB)');

subplot(2,1,2);

plot(F(1:end/2),phaseY(1:end/2));

title('Phase response of signal');

xlabel('Frequency in kHz')

ylabel('radians');

%% Fundamental frequency plot

samples = length(y);

f1 = 0:(fs/samples):(fs/2-(fs/samples)); %find the frequencies of the signal

[Pxx, feq] = pwelch(y, gausswin(fs), fs/2, f1/4, fs);

figure(4)

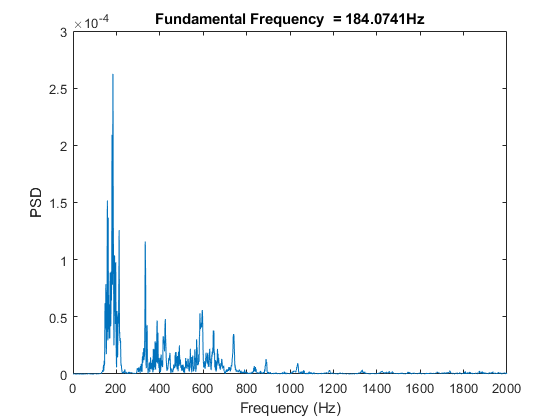
plot(feq,Pxx), ylabel('PSD'), xlabel('Frequency (Hz)');

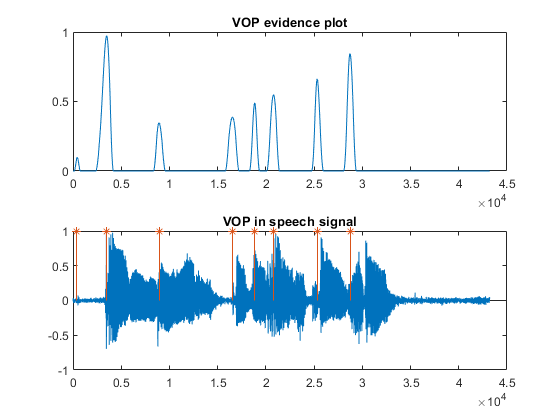
[~, loc] = max(Pxx);

freq\_est = feq(loc);

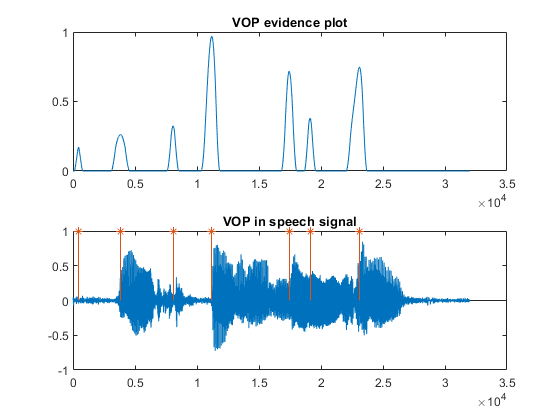
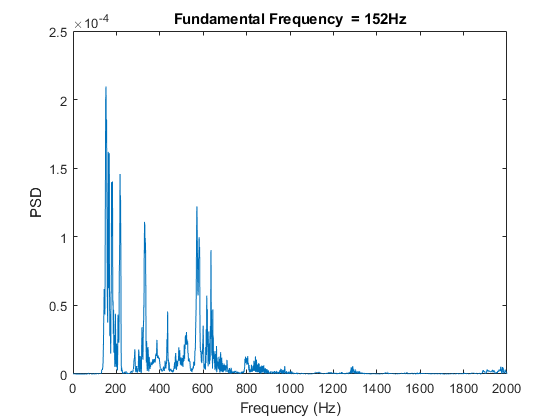
title(['Fundamental Frequency = ', num2str(freq\_est), 'Hz']);

“I\_am\_going\_to\_an\_ice\_cream\_shop.wav”. Output is generated for this file is here





'What\_are\_you\_doing\_here.wav'. Output is generated for this file is here



# References

(Prasanna, Reddy, & Krishnamoorthy, 2009)<https://ieeexplore.ieee.org/document/4802173>

(Vuppala, Yadav, Chakrabarti, & Rao, 2012) <https://ieeexplore.ieee.org/document/6179517>