Speech Signal Processing(EE679) Assignment 1B

Akshay Bajpai(193079002)

27th September 2020

1 Report Structure

For each of the questions , I have written the basic algorithm followed to implement the solution , followed by the code and the output graphs. The codes have been written in octave and the corresponding output files(**plots and .wav files**) are in the data folder submitted .

1.1 Question 1

Use your previous synthesized vowel /u/ at two distinct pitches (F0 = 120 Hz, F0 = 220 Hz). Keep the bandwidths constant at 100 Hz for all formants.

Vowel F1, F2, F3
/u/ 300, 870, 2240

We would like to use the DFT computed with various window lengths and shapes to estimate the vowel's F0 and formant frequencies and study the obtained accuracies with reference to our 'ground truth' values. For the analysis, use a single waveform segment near the centre of your synthesized vowel. Plot the magnitude (dB) spectrum with rectangular and Hamming windows of lengths: 5 ms, 10 ms, 20 ms, 40 ms, each with a large zero-padded DFT.

(i) Comment on the similarities and differences between the different computed spectra.
(ii) Estimate the signal parameters from each of the magnitude spectra and report the error with respect to the ground-truth

Solution :
Algorithm :

1. We use the code segment used in the earlier assignment to create a filter corresponding to the formant frequencies 300Hz,870Hz and 2240Hz.
2. For each of the formant frequencies , we obtain the values of r and θ . Equations used are as below:

$$r_1 = exp(\frac{-B_1\pi}{F_s}) \tag{1}$$

$$\theta_1 = \frac{2\pi F_1}{F_s} \tag{2}$$

3. The values obtained above are used to calculate the frequency response equation coefficients.

$$H(z) = \frac{1}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \tag{3}$$

4. Three pairs of the coefficients obtained, corresponding to each formant frequency, are convolved to obtain a combined frequency response for the vowel u .

5. Triangular pulse train corresponding to F_0 $120Hz$ and $220Hz$ are approximated using the square function and input to the filter.

6. The output of the filter is obtained using the difference equation below:

$$y(n) = b_0 x(n) - a_1 y(n-1) - a_2 y(n-2) \tag{4}$$

7. We generate hamming and rectangular of window of a given length , multiply the output signal $y(n)$ and compute the 1024 point DFT .
   Hamming Window:
   $$w(n) = 0.54 + 0.46cos(\frac{2\pi n}{N-1}) \tag{5}$$
   Rectangle Window:
   $$w(n) = \begin{cases} 1, & \text{if } 0 \leq x \leq N \\ 0, & \text{otherwise} \end{cases}$$

8. The length of the window is found using the relation below. Time duration of the window multiplied with the sampling frequency.

$$N = (Time_{duration})F_s \tag{6}$$

Code :

Main Script :

```
pkg load signal;

f1=300;
f2=870;
f3=2240;
filename="u";
b0=100;
fs=16000;
f0=[120,220];

for j=1:columns(f0)
    r1 = exp(-b0*pi*1/fs);
    theta1 = 2*pi*f1*1/fs;
    r2 = exp(-b0*pi*1/fs);
    theta2 = 2*pi*f2*1/fs;
    r3 = exp(-b0*pi*1/fs);
    theta3 = 2*pi*f3*1/fs;
```

```
    poles1 = [r1*exp(1j*theta1) , r1*exp(-1j*theta1)];
    poles2 = [r2*exp(1j*theta2) , r2*exp(-1j*theta2)];
    poles3 = [r3*exp(1j*theta3) , r3*exp(-1j*theta3)];

    b1 = [1  ,0,0];
    a1= [1,-2*r1*cos(theta1),r1**2];
    b2 = [1  ,0,0];
    a2= [1,-2*r2*cos(theta2),r2**2];
    b3 = [1  ,0,0];
    a3= [1,-2*r3*cos(theta3),r3**2];
    b_temp=conv(b1,b2);
    b = conv(b_temp,b3);
    a_temp=conv(a1,a2);
    a = conv(a_temp,a3);
    [h,w] = freqz(b,a) ;

    k=figure;
    plot(fs*w/(2*pi),20*log10(abs(h)));
    xlabel('Frequency (Hz)');
    ylabel('Magnitude (dB)');
    title(['Frequency response for filter /',filename,'/']);
    grid on;
    saveas(k,sprintf('output/Frequency_response_%s.png',filename));
    y=input_signal(h,b,a,f0(1,j),fs,0.5,filename);
    windowing(y,w,fs,f0(1,j));
endfor
```

Function to calculate filter output :

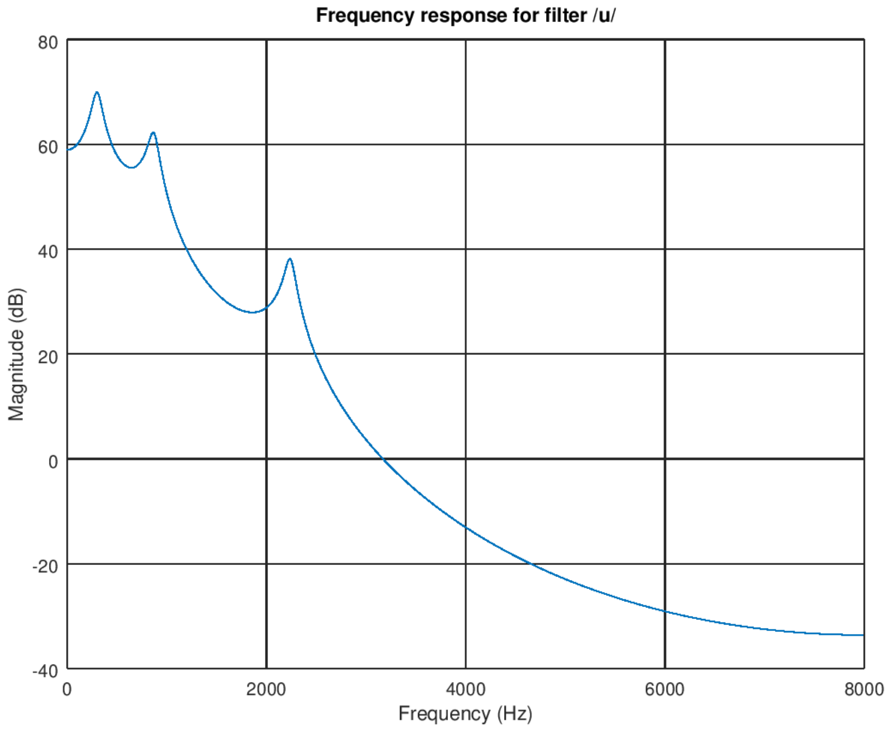
```
  function y=input_signal(h,b,a,f0,fs,time_,filename)

  t = 0:time_/fs:time_; % 0s to 0.5s with Fs sample freq
  [z,time_length]=size(t);
  x1 = max(0,(square(2*pi*f0*t,0.01)));  % 2*pi* freq * time duration
  x1=x1';
  y=zeros(time_length,1);

  [ro,col]=size(y);
  [m,number_of_poles]=size(a);

  for i=number_of_poles:rows(y)
    y(i,1) = b(1,1)*x1(i-2,1) ;
    for j=2:number_of_poles
      y(i,1)=y(i,1)-a(1,j)*y(i-j+1,1);
    end
  end

  k=figure;
  t=t';
```

```
  plot(t,y);
  xlabel('Time(s)');
  ylabel('Magnitude');
  title(['Filter output for formants corresponding to /',filename,'/ and F0 =', num2str(f0)]);
  grid on;
  saveas(k,sprintf('output/Filter_output_F0_%d.png',f0));
  close(k);
  wavwrite (y,fs, ["output/",filename,"_",num2str(f0),".wav"]);
```

Windowing function :

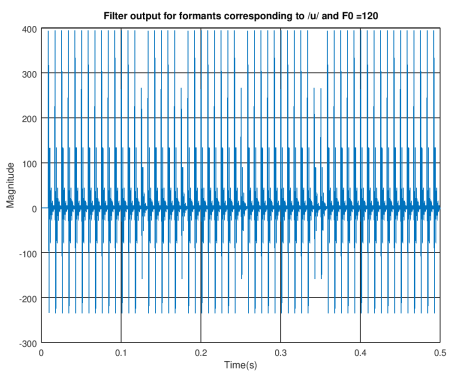
```
   function windowing(y,om,fs,f0)
 m=[0.005,0.010,0.020,0.040];

for i=1:columns(m)
  time_=m(1,i);
  M = time_ * fs;
  r=y(8000:8000+M,1);
  w = .54 + .46*cos(pi*(-M/2:M/2)/M);

  result = r'.*w;
  display(size(result));
  result = fft(result,1024);
  k=figure;
  f = fs*(1:(512))/1024;
  plot(f,20*log10(abs(result(1:length(result)/2))));
  xlabel('Frequency(Hz)');
  ylabel('Magnitude (dB)');
  title(['Hamming Window output for F0 =', num2str(f0),' and window length =',num2str(m(1,i)),'s']);
  grid on;
  saveas(k,sprintf('output/hamming/Hamming_F0_%d_%fs.png',f0,m(1,i)));
  close(k);
endfor

for i=1:columns(m)
  time_=m(1,i);
  M = time_ * fs;
  r=y(8000:8000+M,1);
  w = ones(1:M,1);
  result = r'.*w;
  display(size(result));
  result = fft(result,1024);
  k=figure;
  f = fs*(1:(512))/1024;
  plot(f,20*log10(abs(result(1:length(result)/2))));
  xlabel('Frequency(Hz)');
  ylabel('Magnitude (dB)');
  title(['Rectangular Window output for F0 =', num2str(f0),' and window
```

```
        length=',num2str(m(1,i)),'s']);
    grid on;
    saveas(k,sprintf('output/rectangle/Rectangular_F0_%d_%fs.png',f0,m(1,i))
        );
    close(k);

endfor
```

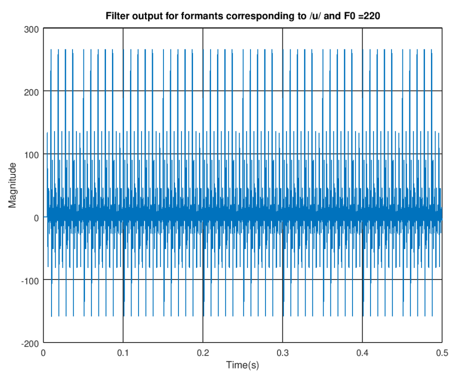


(a)

Figure 1: (a)Frequency response of the filter for vowel u



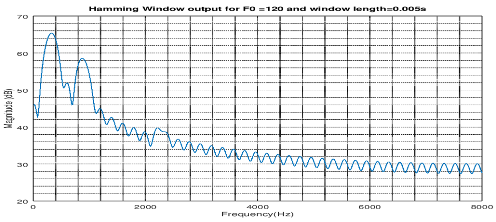
(a)



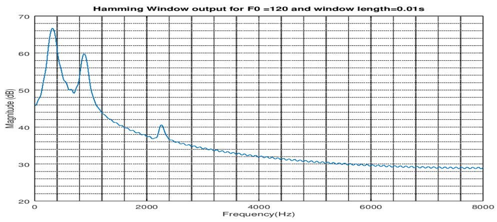
(b)

Figure 2: (a)Response of the filter of $\mathrm{F}_0 = 120\mathrm{Hz}(b)Responseofthefilterof F_0 = 220\mathrm{Hz}$

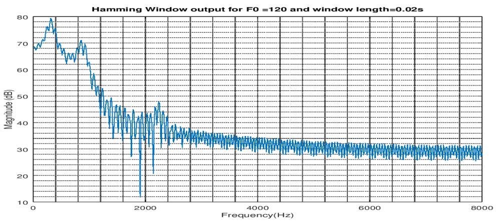
Hamming Window outputs:



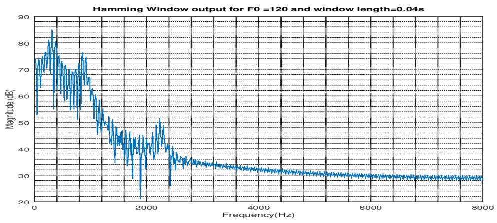
(a) Hamming window 5ms



(b) Hamming window 10ms



(c) Hamming window 20ms

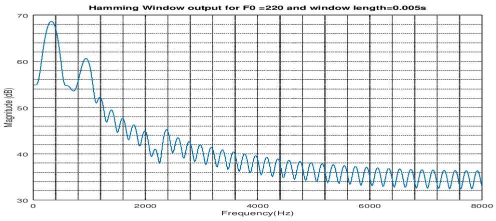


(d) Hamming window 30ms

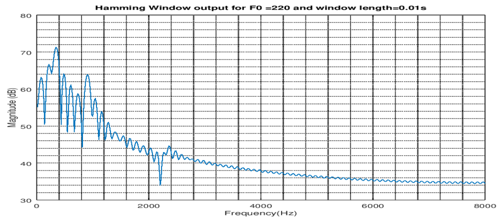
Figure 3: Hamming window response for $F_0 = 120Hz$

Manual Observations from the plot Hamming window $F_0 = 120Hz$						
	Observed Values			Errors		
Length	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)	Formant 1(340 Hz)	Formant 2(870 Hz)	Formant 3(2240 Hz)
5ms	335	880	2234	5	10	6
10ms	335	888	2271	5	18	31
20ms	317	851	2252	23	19	12
40ms	335	870	2234	5	0	6

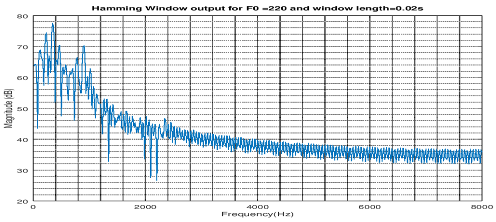
Table 1: Observations from the Hamming window Plots $F_0 = 120Hz$



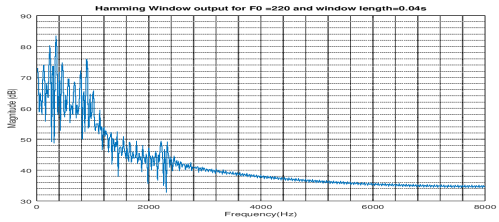
(a) Hamming window 5ms



(b) Hamming window 10ms



(c) Hamming window 20ms



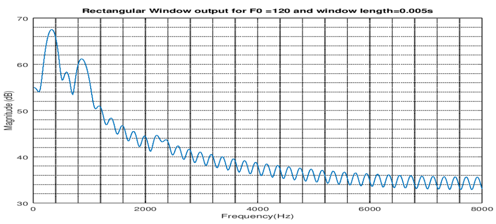
(d) Hamming window 30ms

Figure 4: Hamming window response for $F_0 = 220Hz$

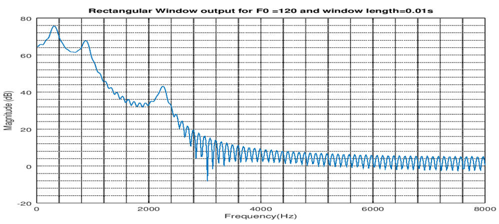
Manual Observations from the plot Hamming window $F_0 = 220Hz$						
Observed Values				Errors		
Length	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)	Formant 1(340 Hz)	Formant 2(870 Hz)	Formant 3(2240 Hz)
5ms	335	943	2380	5	103	140
10ms	353	925	2380	13	85	140
20ms	353	906	2344	13	66	104
40ms	335	888	2215	5	48	25

Table 2: Observations from the Hamming window Plots $F_0 = 220Hz$

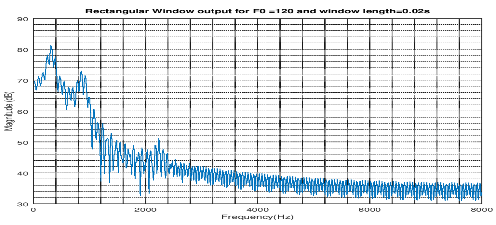
Rectangular Window Outputs:



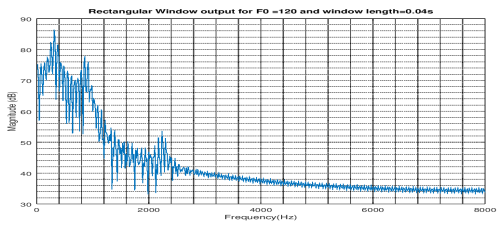
(a) Rectangular window 5ms



(b) Rectangular window 10ms



(c) Rectangular window 20ms

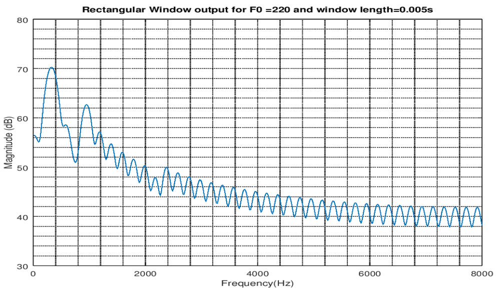


(d) Rectangular window 30ms

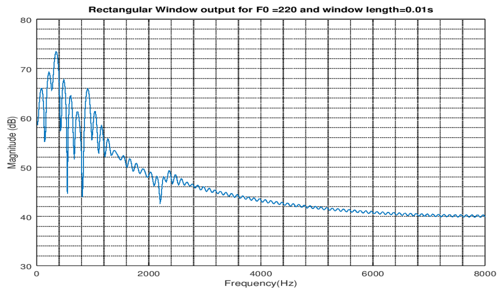
Figure 5: Rectangular window response for $F_0 = 120Hz$

Manual Observations from the plot Rectangular window $F_0 = 120Hz$						
Observed Values				Errors		
Length	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)	Formant 1(340 Hz)	Formant 2(870 Hz)	Formant 3(2240 Hz)
5ms	335	860	2215	5	10	25
10ms	335	888	2260	5	18	20
20ms	335	870	2252	5	0	12
40ms	335	870.05	2245	5	0.05	5

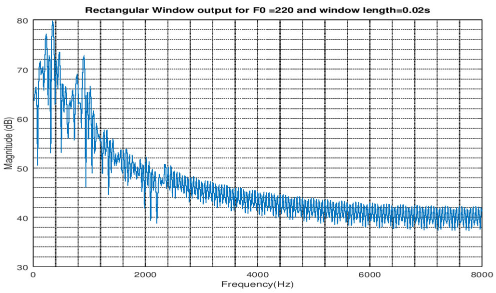
Table 3: Observations from the Rectangular window Plots $F_0 = 120Hz$



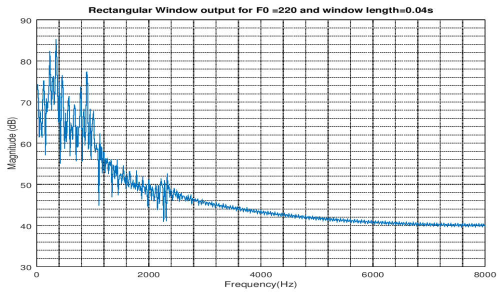
(a) Rectangular window 5ms



(b) Rectangular window 10ms



(c) Rectangular window 20ms



(d) Rectangular window 30ms

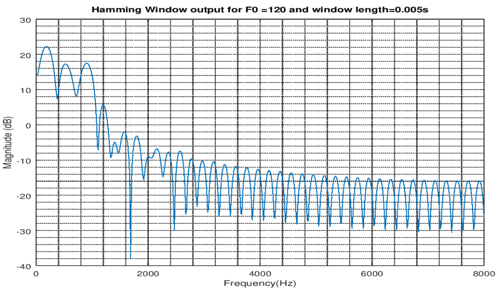
Figure 6: Rectangular window response for $F_0 = 220Hz$

Manual Observations from the plot Rectangular window $F_0 = 220Hz$						
Observed Values				Errors		
Length	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)	Formant 1(340 Hz)	Formant 2(870 Hz)	Formant 3(2240 Hz)
5ms	335	943	2381	23	103	141
10ms	353	906	2363	13	66	123
20ms	353	906	Not Clear	13	66	–
40ms	353	888	Not Clear	13	48	–

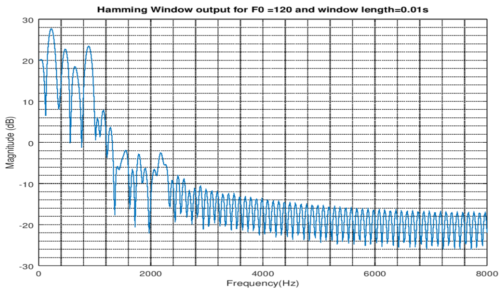
Table 4: Observations from the Rectangular window Plots $F_0 = 220Hz$

1.2 Output from the /u/ waveform given on moodle:

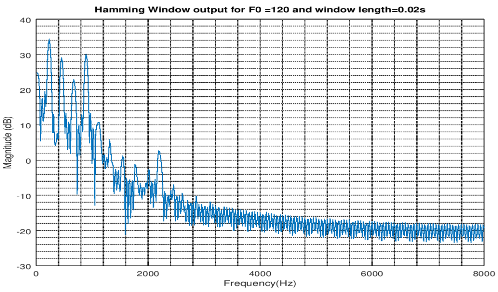
Hamming Window outputs:



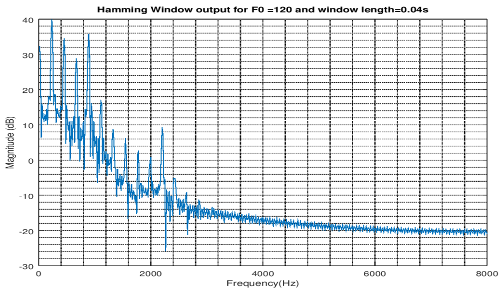
(a) Hamming window 5ms



(b) Hamming window 10ms

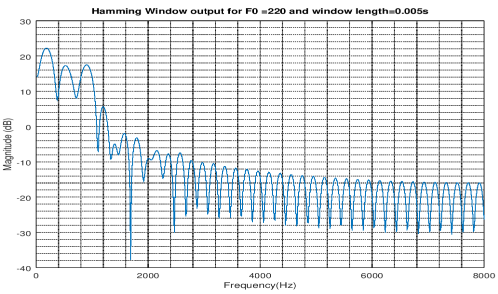


(c) Hamming window 20ms

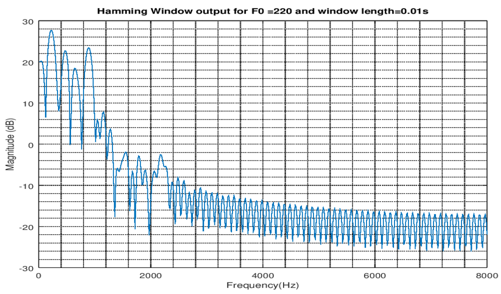


(d) Hamming window 30ms

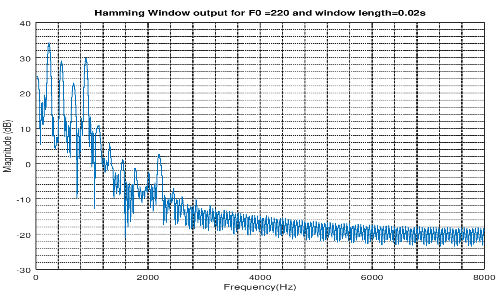
Figure 7: Hamming window response for $\mathrm{F}_0 = 120Hz$



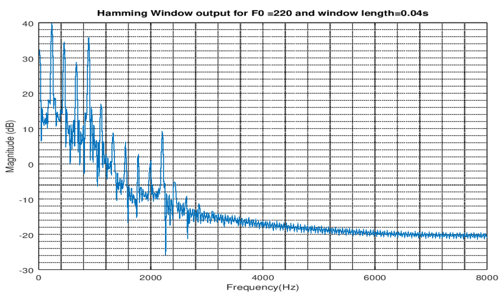
(a) Hamming window 5ms



(b) Hamming window 10ms



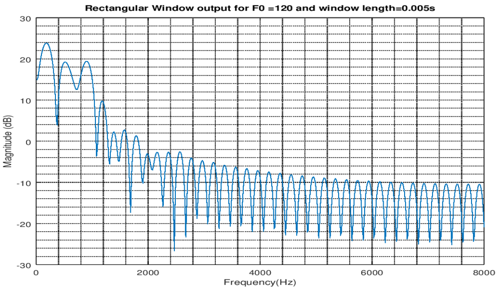
(c) Hamming window 20ms



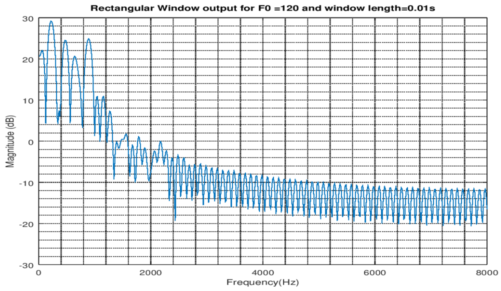
(d) Hamming window 30ms

Figure 8: Hamming window response for $\mathrm{F}_0 = 220Hz$

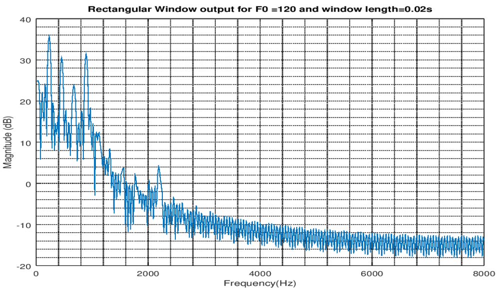
Rectangular Window outputs:



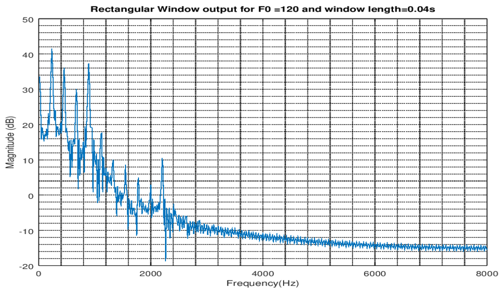
(a) Rectangular window 5ms



(b) Rectangular window 10ms

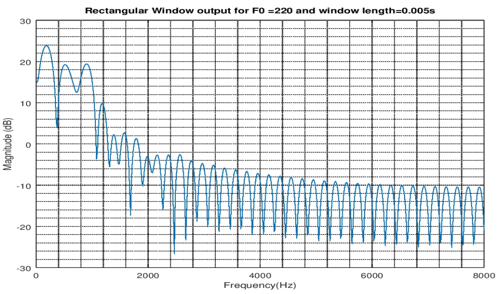


(c) Rectangular window 20ms

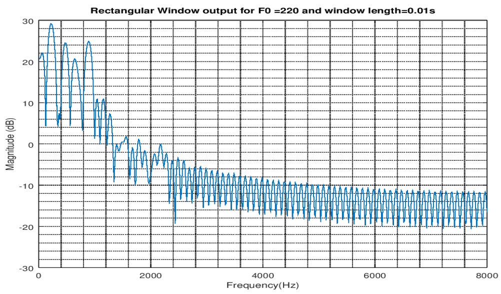


(d) Rectangular window 30ms

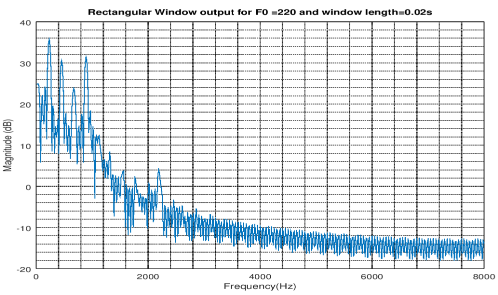
Figure 9: Rectangular window response for $F_0 = 120Hz$



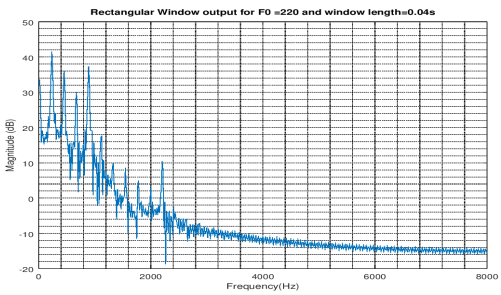
(a) Rectangular window 5ms



(b) Rectangular window 10ms



(c) Rectangular window 20ms



(d) Rectangular window 30ms

Figure 10: Rectangular window response for $F_0 = 220Hz$

1.3 Comments:

1. Both in the case of rectangular and hamming window, as we increase the window duration, we begin to see more harmonics in the plot.
2. This is expected, since with increase in the window duration, the output is more narrow band.
3. For a small window duration 5 ms , the output is a smooth plot , where different harmonics have merged to give a smooth output.
4. With increase in the fundamental frequencies , F_0, the output seems to be more noisy as more harmonics are present.
5. For the case of rectangular window, the side lobe have a higher energy as compared to hamming window of same duration. This is expected as in the case of hamming window, the side lobe energy is 40 dB less than the main lobe.