EE679: Speech Processing

Computing Assignment 1

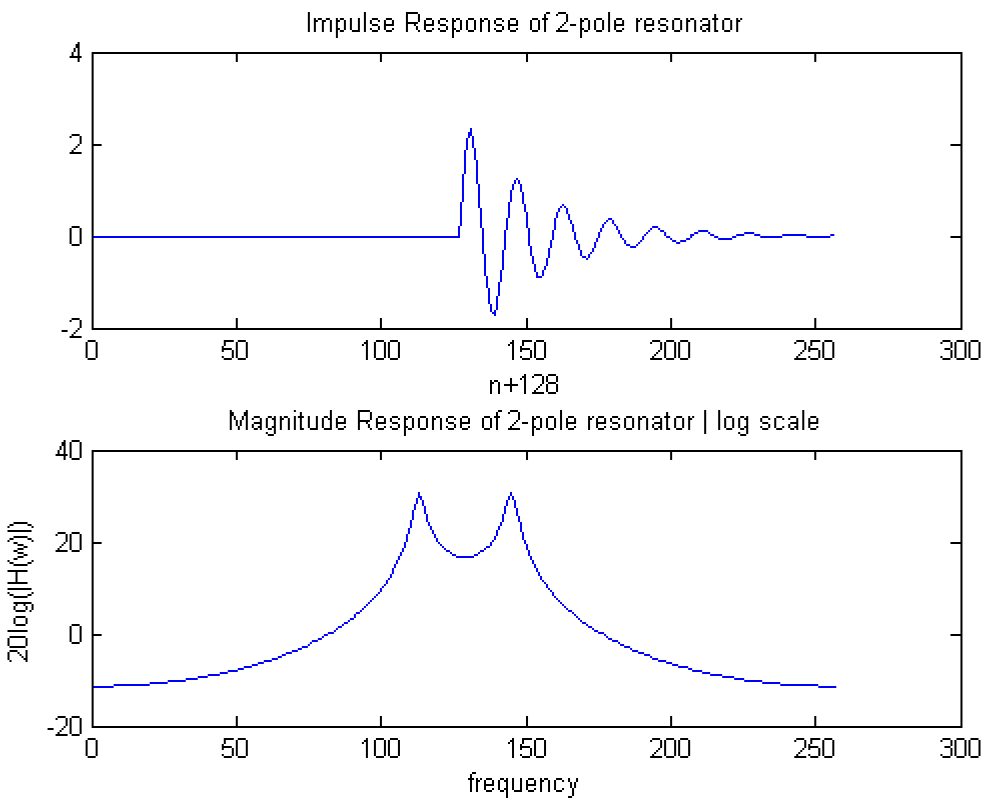
Ashwin Kachhara
10d070048

**Q1**

Given that F1 = 1 kHz,
B1 = 200Hz,
Fsampling = 16 kHz

We are synthesizing a 2 pole digital resonator with given formant.
From the values of F1 and B1, we can calculate the locations of the poles of the filter ($re^{j\theta}$ and $re^{-j\theta}$): as follows:
$r = e^{-\pi*B1*T}$ $\theta = 2*\pi*F1*T$, where T = 1/Fs

Hence we can find the transfer function H(z) of the filter. On substituting $z = e^{jw}$, into the above, we get the fourier transform H(w) of the filter. Taking absolute value of that gives us the magnitude response of the filter




.

Now, we have to obtain the impulse response of the function from the transfer function.

To do that, we obtain the difference equation corresponding to the transfer function H(z), which turns out to be the following.

$y[n] - 2r\cos\theta * y[n-1] + r^2 y[n-2] = x[n]$

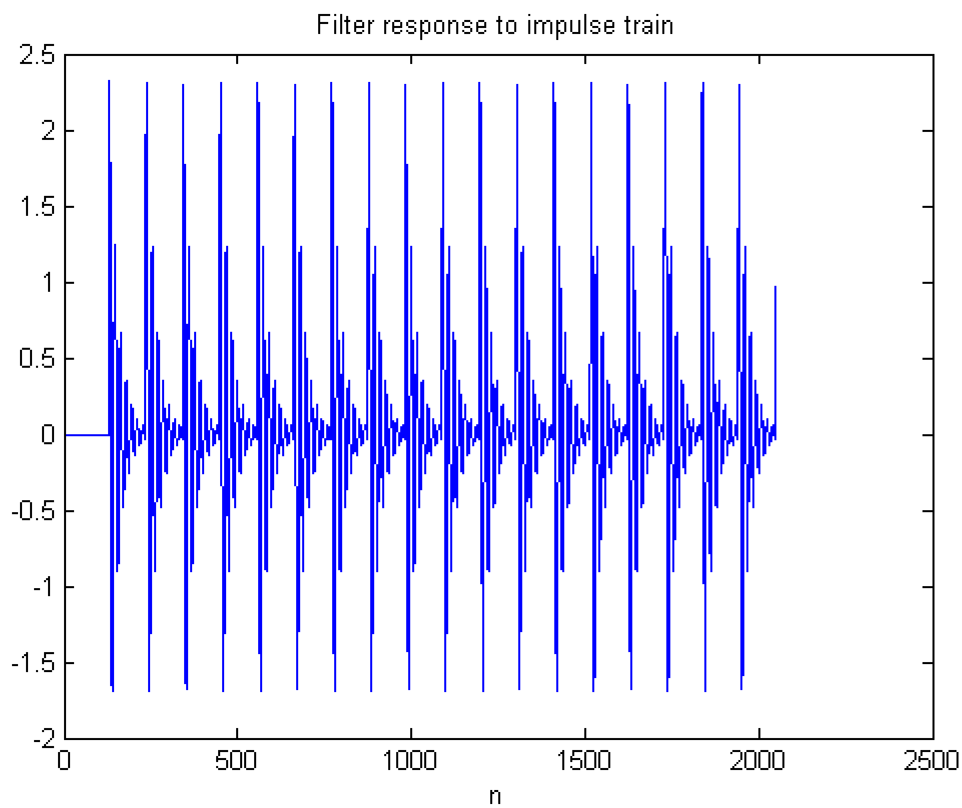
So, if we set $x[n] = \delta[n]$
Then evaluating $y[n] = h[n]$

On evaluating y[n] on 256 points on the unit circle, we get the graph of h[n] as shown above

**<u>Q2</u>**

Excitation of the filter is done by an impulse train of F0 = 150 Hz
Defining the impulse train is quite easy.

The output obtained on passing the impulse train through the filter is as shown below:

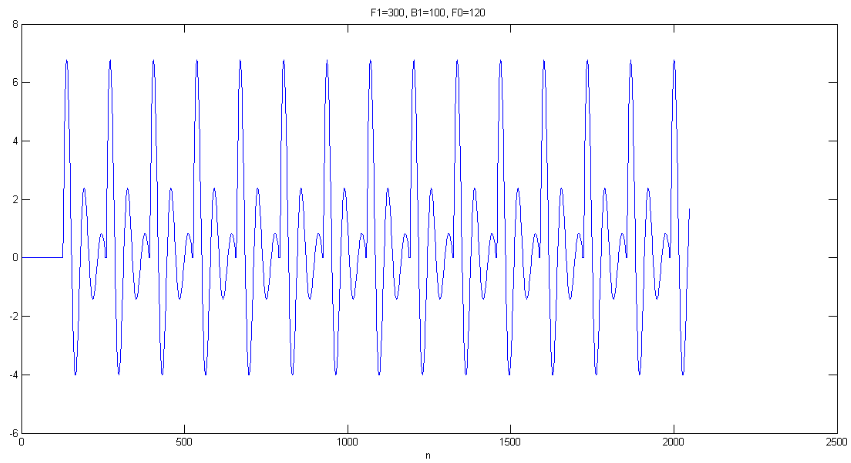



As expected it is the same filter impulse response repeated for each peak in the impulse train.

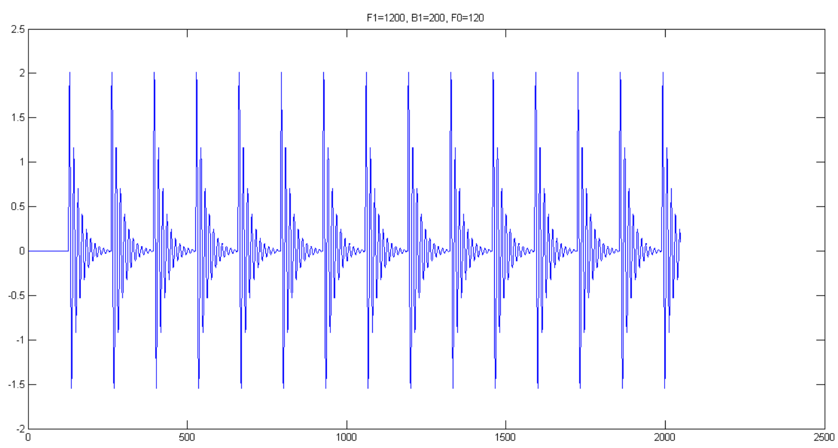
The sound quality is quite sharp and it sounds very artificial.

**Q3**

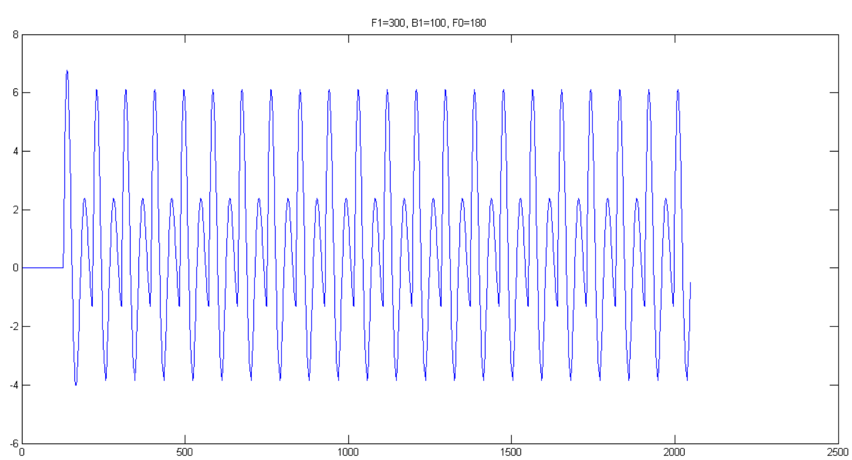
(a) F1 = 300 Hz, B1 = 100 Hz, F0 = 120 Hz



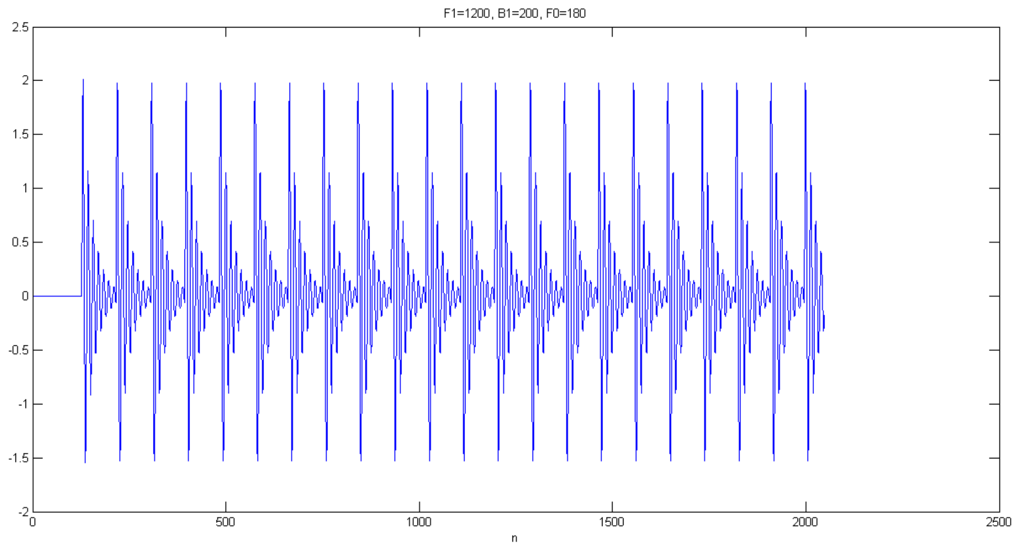

(b) F1 = 1200 Hz, B1 = 200 Hz, F0 = 120 Hz




(c) F1 = 300 Hz, B1 = 100 Hz, F0 = 180 Hz



(d) F1 = 1200 Hz, B1 = 200 Hz, F0 = 180 Hz




Comparison:

The filters in (b) and (d) are of higher formant frequency, which is visible from the waveform. The peaks are much closer together. The filters of (a) and (c) are of much lower formant frequency.

The excitation impulse train is of higher frequency in (c) and (d). The impulse train peaks are closer together in time. Hence the filter impulse response has had shorter time to decay.

Comparing the sounds of (a) and (c), the latter sounds higher pitched compared to the former due to the higher excitation frequency, which is closer to the filter resonance frequency.

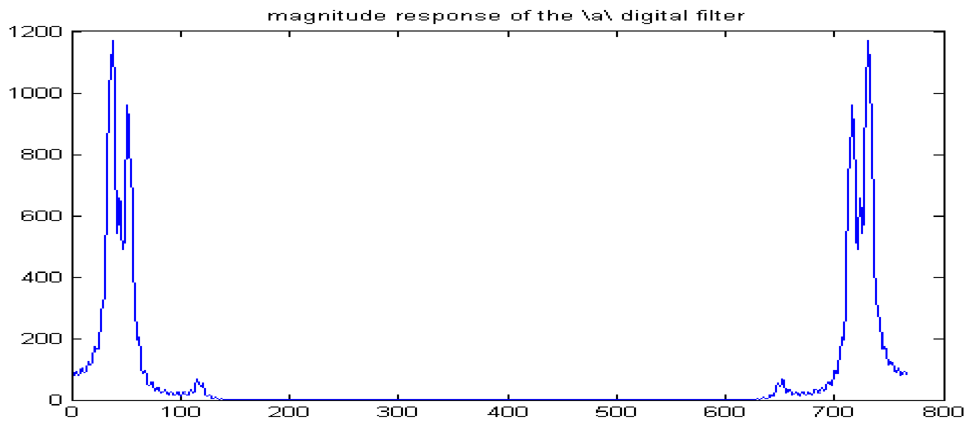
On, comparing the sounds of (b) and (d) both of these sound higher compared to the previous two, because this particular filter has resonance at 1200 Hz as compared to 300 Hz in the previous. Of course the latter sounds higher pitched compared to the former due to higher excitation frequency.

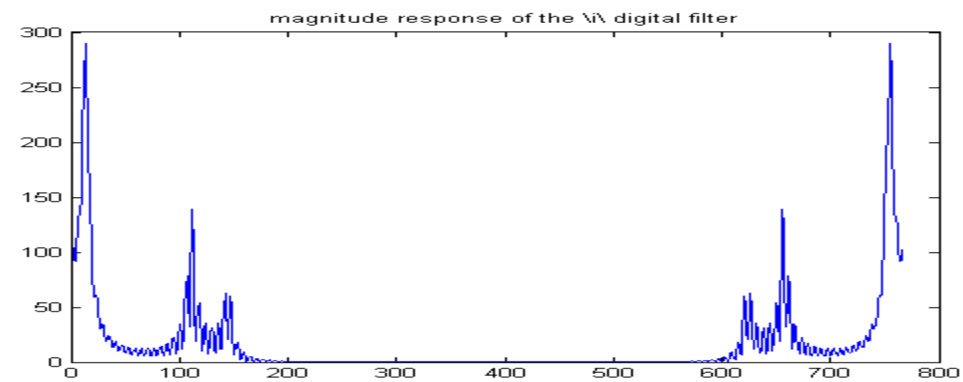
**Q4**

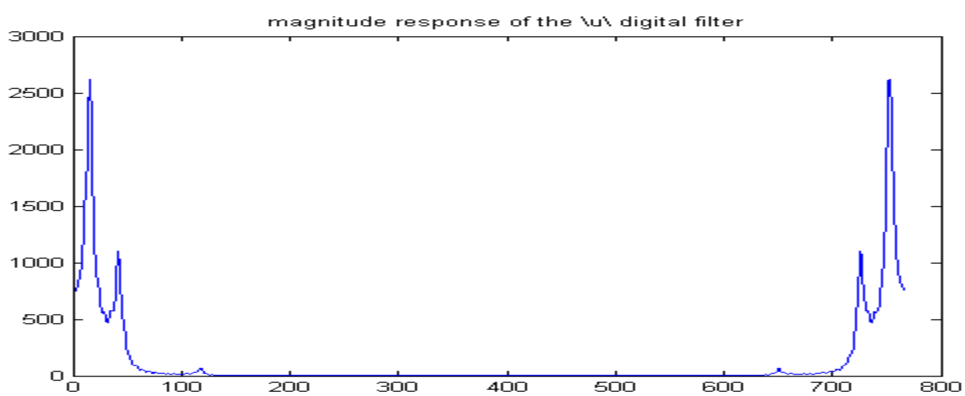
We can get the transfer function of the 3 formant filter by simply multiplying the transfer functions of single formant filters, each of whose frequency is a formant of our required filter.

H(z) = H1(z)H2(z)H3(z)

Vowel F1, F2, F3
/a/ 730, 1090, 2440
/i/ 270, 2290, 3010
/u/ 300, 870, 2240
So, we get the following magnitude responses for the three vowel filters:

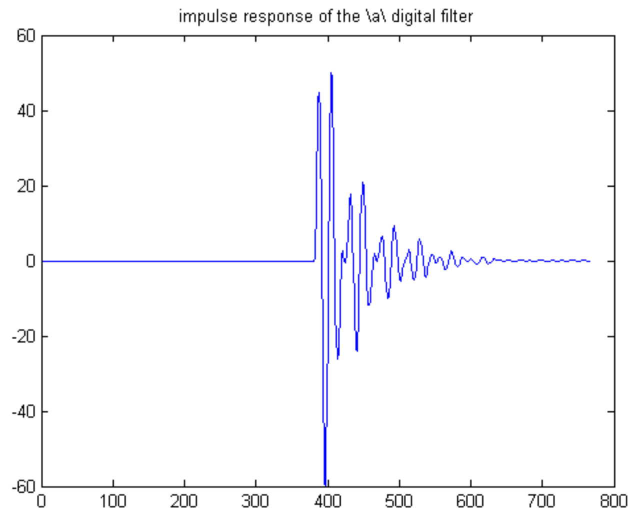


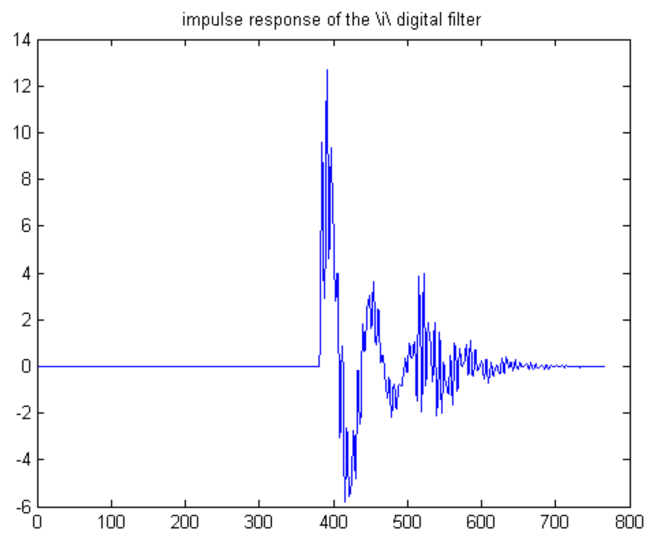



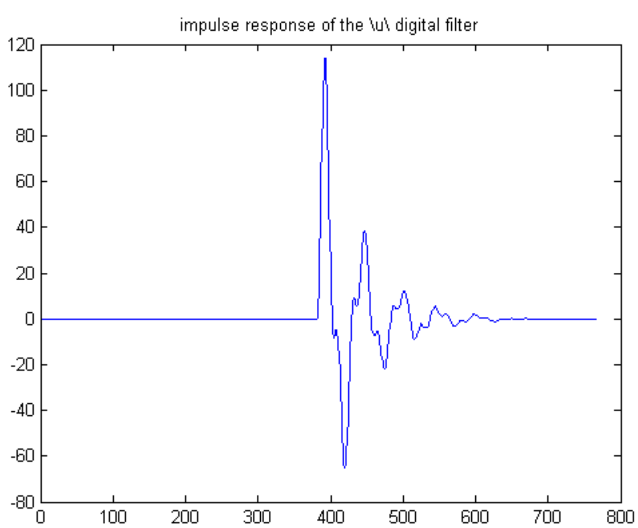

The above figures represent the magnitude response of each filter from [0, 2π]

Impulse response of each filter

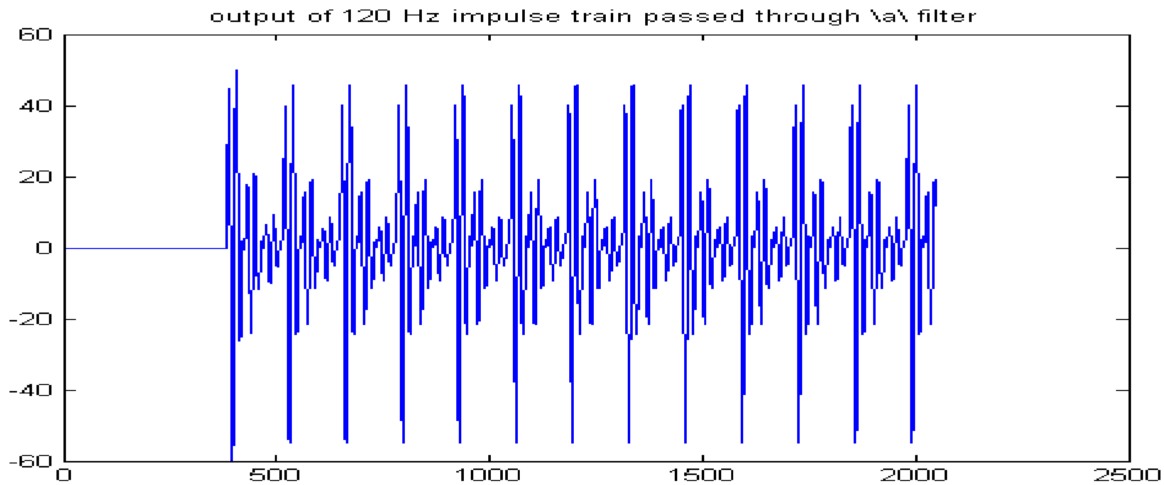

impulse response of the \a\ digital filter
60
40
20
0
-20
-40
-60
0
100
200
300
400
500
600
700
800

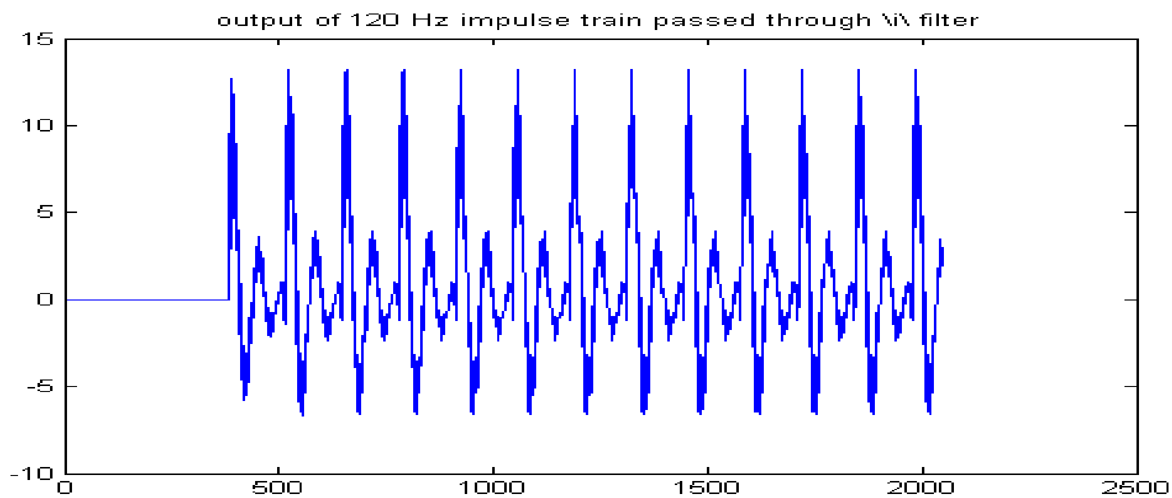
:

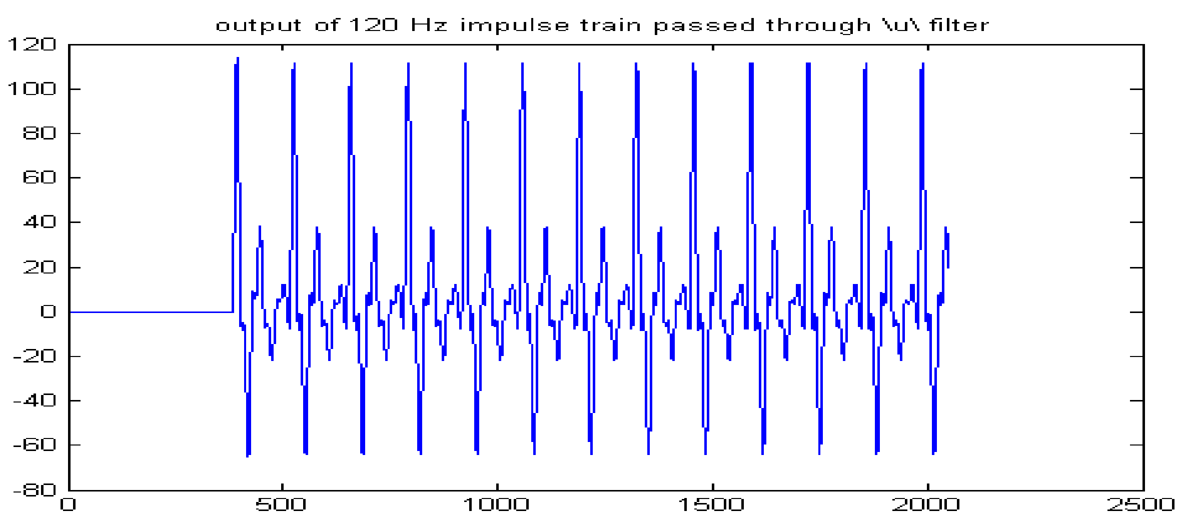

impulse response of the \i\ digital filter
14
12
10
8
6
4
2
0
-2
-4
-6
0
100
200
300
400
500
600
700
800


impulse response of the \u\ digital filter
120
100
80
60
40
20
0
-20
-40
-60
-80
0
100
200
300
400
500
600
700
800

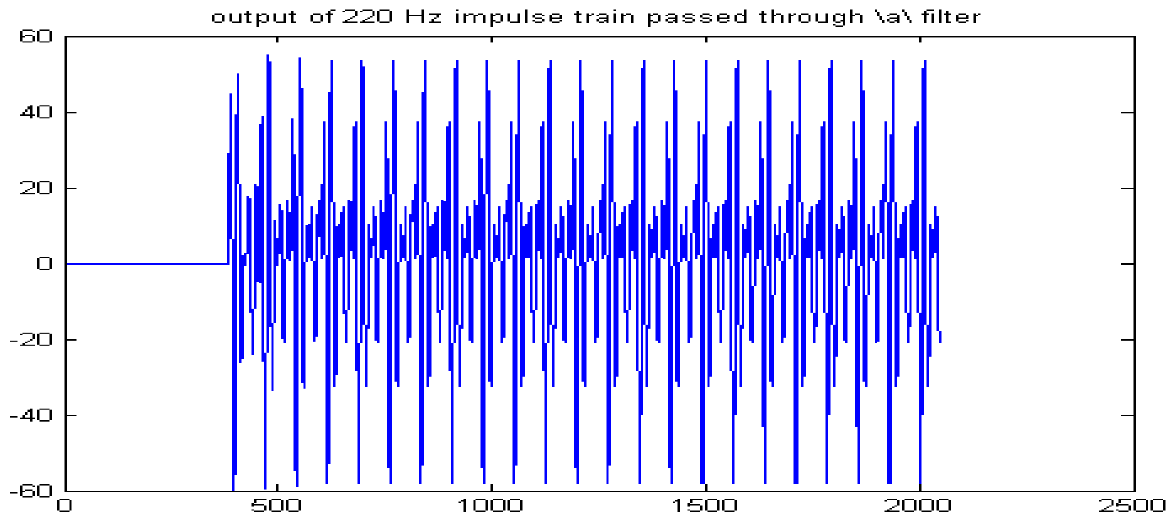
The resultant waveforms from passing the impulse trains through these filters are as follows:
For impulse train of 120 Hz:

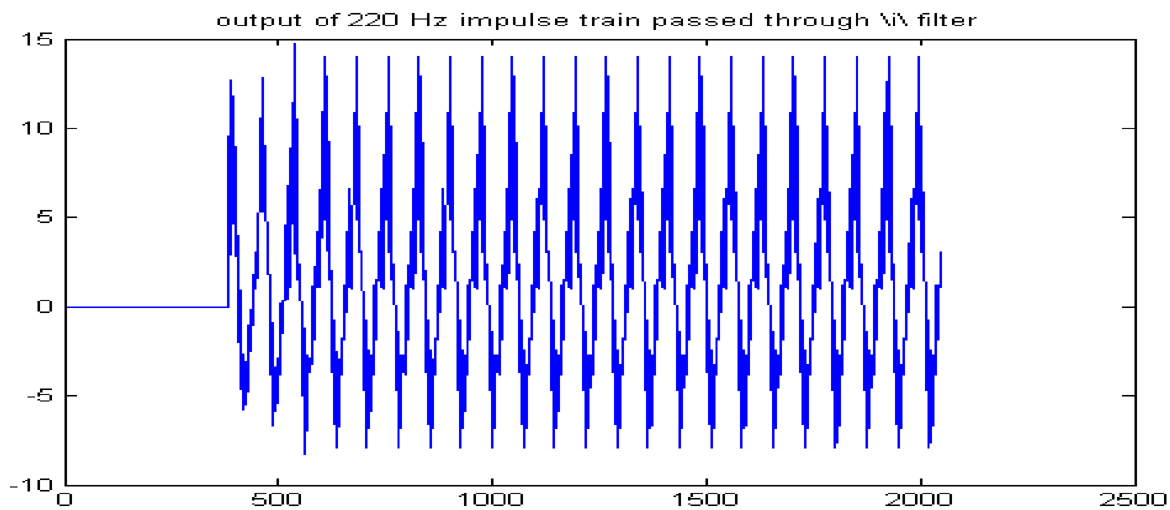


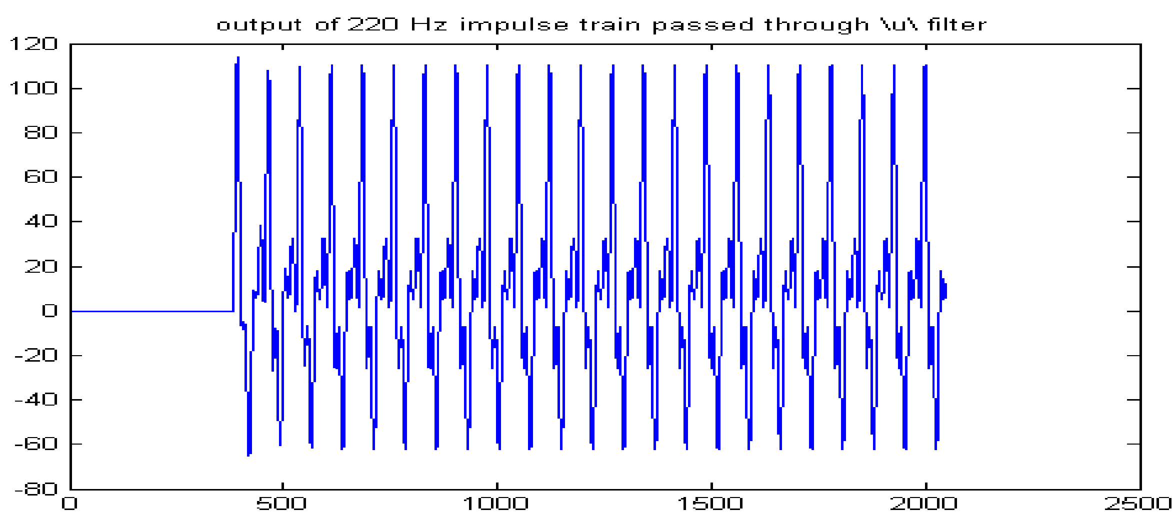



For 220 Hz impulse train:



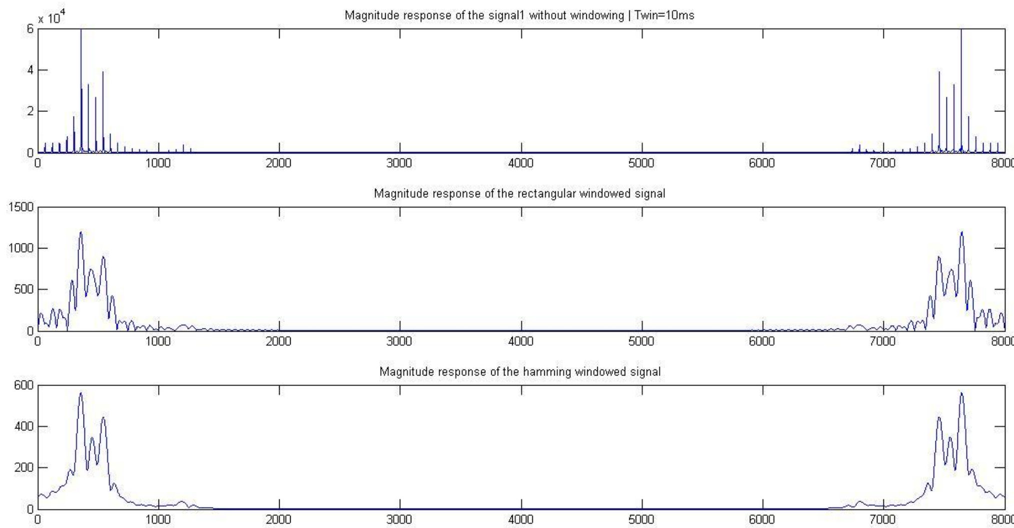



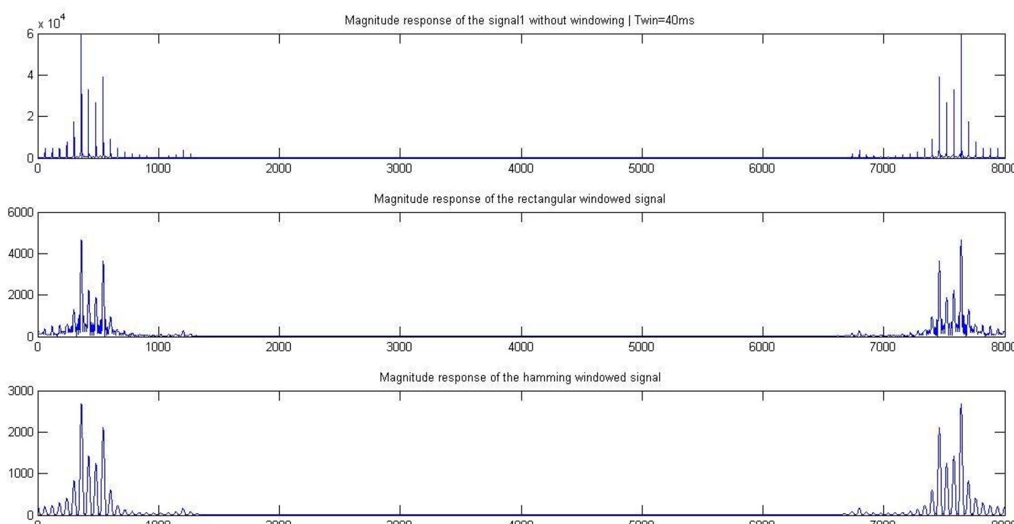
Sampling frequency in each case was 16000 Hz

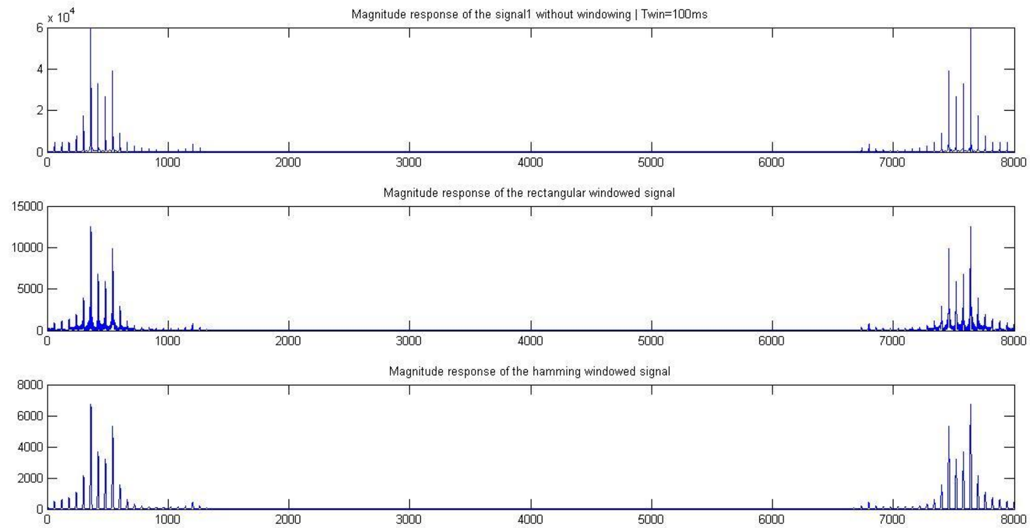
## Q5

Signal1 = ya120; Signal2 = yi220

For signal1:
Figures of rectangular and hamming windows of lengths 10ms, 40ms and 100ms and magnitude response of signal without use of any window.






For all the figures [0, 8000] => [0, 2π]

<u>Observations:</u> The peaks peaks appear to be enhanced on use of longer window. For shorter window lengths, it may be harder to get an accurate estimate of the signal parameters, while the parameters can be more easily extracted by using a longer window. In general, the hamming window gives better performance because the side lobes decay very fast, so it causes less interference with the impulses in the neighborhood.

<u>Parameter Extraction:</u> We will use the Twin = 100ms hamming windowed function for extraction of parameters.
From the we see 3 peaks at approximately n=6800, 7500, 7650
Relation b/n frequency and n is
$F_i = (8000 - n_i)*Fs/8000$; $Fs = 16000$
So, the formants of this signal are at approximately 2400 Hz, 1000 Hz and 700 Hz

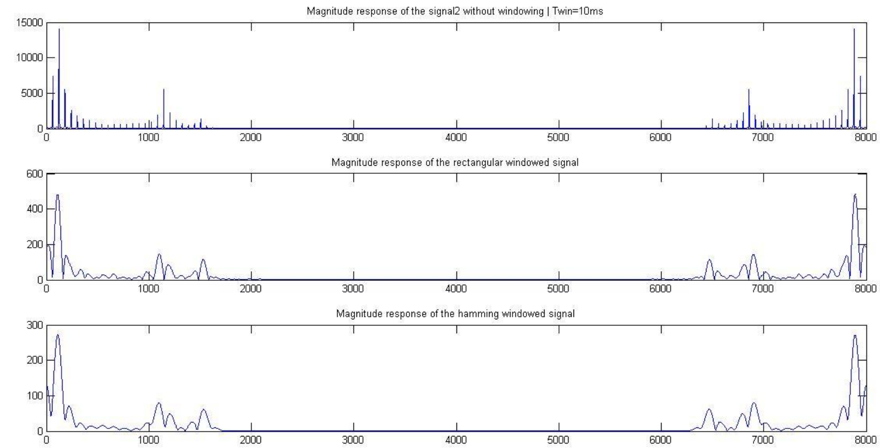
The pitch of the signal can be extracted from the unwindowed signal using the spacing between the impulses. In this case, the spacing is slightly more than 50 samples. With a sampling frequency of 16 kHz, the spacing corresponds to > 50*16000/8000
Therefore the pitch corresponds to slightly more than 100 Hz.

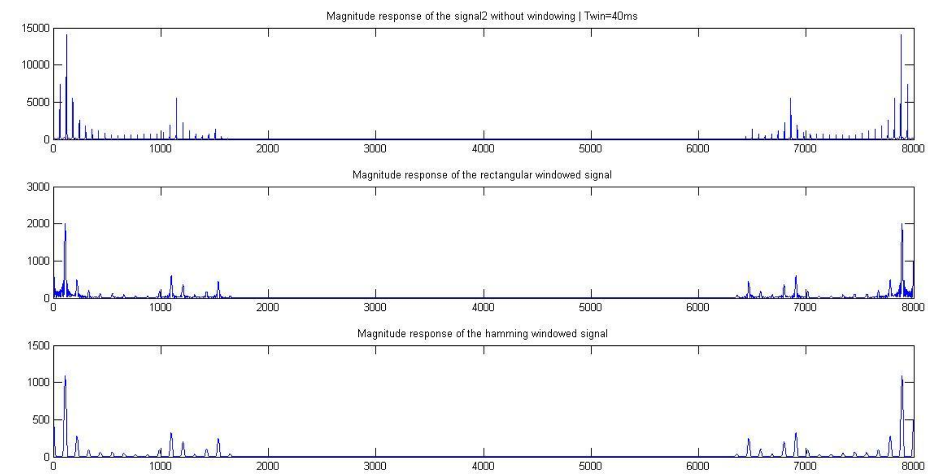
Now, the original parameters were:
Formants at 730 Hz, 1090 Hz, and 2440 Hz
Pitch was 120 Hz

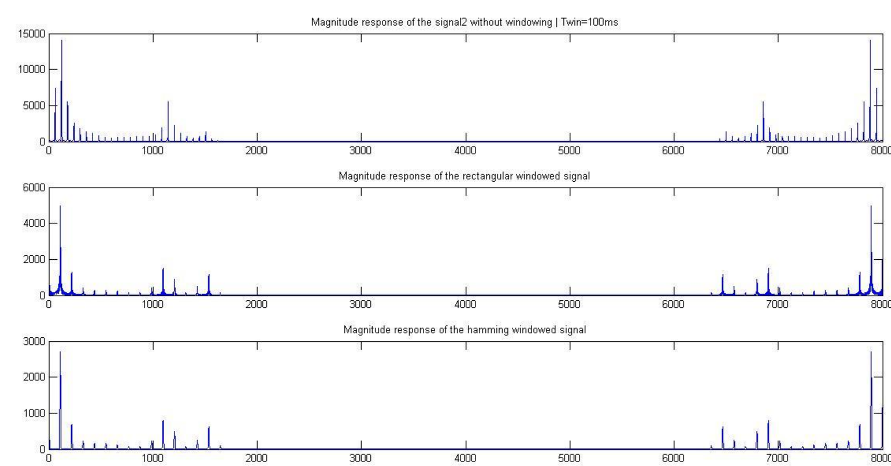
We have gotten a very reasonable estimate of the signal parameters using this analysis.

For signal 2:

Figures of rectangular and hamming windows of lengths 10ms, 40ms and 100ms and magnitude response of signal without use of any window.




For all the figures [0, 8000] => [0, 2π]

Observations: The peaks peaks appear to be enhanced on use of longer window. For shorter window lengths, it may be harder to get an accurate estimate of the signal parameters, while the parameters can be more easily extracted by using a longer window. In general, the hamming window gives better performance because the side lobes decay very fast, so it causes less interference with the impulses in the neighborhood.

Parameter Extraction: We will use the Twin = 100ms hamming windowed function for extraction of parameters.
From the we see 3 peaks at approximately n=6500, 6900, 7900
Relation b/n frequency and n is
$F_i = (8000 - n_i)*Fs/8000$; $Fs = 16000$
So, the formants of this signal are at approximately 3000 Hz, 2200 Hz and 200 Hz

The pitch of the signal can be extracted from the unwindowed signal using the spacing between the impulses. In this case, the spacing is slightly more than 70 samples. With a sampling frequency of 16 kHz, the spacing corresponds to > 70*16000/8000
Therefore the pitch corresponds to slightly more than 140 Hz.

Now, the original parameters were:
Formants at 270 Hz, 2290 Hz, and 3010 Hz
Pitch was 220 Hz

We have gotten a very reasonable estimate of the signal parameters using this analysis.