

Speech Processing EE679

Mohit
20D070052

August 2023



1 Student Details

Name: Mohit
Roll No: 20D070052

2 Question 1

Given the following specification for a single-formant resonator, obtain the transfer function of the filter $H(z)$ from the relation between resonance frequency / bandwidth, and the pole angle / radius. Plot filter magnitude response (dB magnitude versus frequency) and impulse response.

F1 (formant) = 900 Hz

B1 (bandwidth) = 200 Hz

Fs (sampling freq) = 16 kHz

2.1 Solution 1

We know that the Vocal Tract transfer function $H(z)$, for given formant frequency F_i and bandwidth B_i is given by the formula:

$$H(z) = \left\{ \frac{1}{(1-re^{-j\theta}z^{-1})(1-re^{j\theta}z^{-1})} \right. \\ \left. \frac{1}{1-2r\cos(\theta)z^{-1}+r^2z^{-2}} \right\}$$

where, $r = e^{-\sigma_p T_s}$ and $\theta = \Omega_p * T_s$
and $\sigma_p = \pi B_i$ and $\Omega_p = F_i$

The Impulse Response of the filter is calculated using the difference equation formula:

Summation $k=0$ to M $b[k]y[n-k] = \text{Summation } k=0 \text{ to } N \text{ } a[k]x[n-k]$

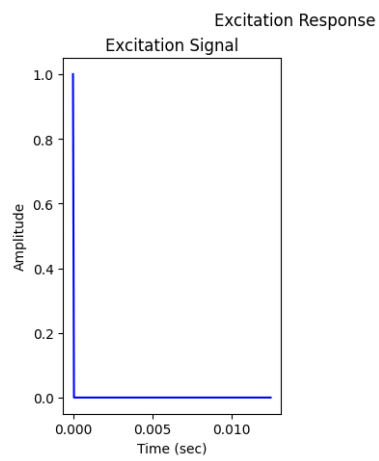
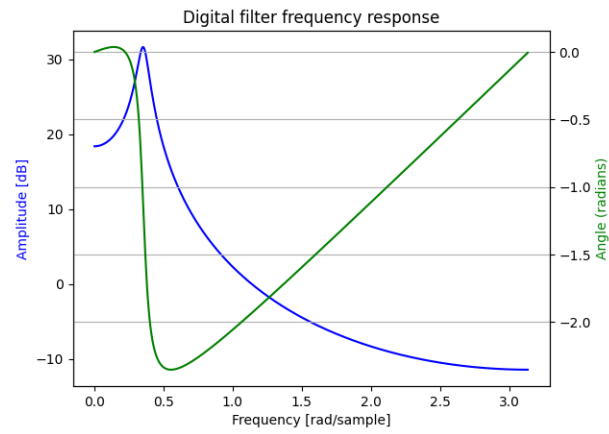
where, $y[n]$ is the Impulse Response, $b[k], a[k]$ are the numerator and denominator coefficients of $H(z)$ and $x[n]$ is the input Impulse signal

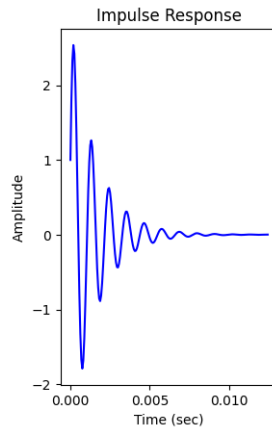
Implementing both of these into codes we get the following results.

2.2 Code

The code files are included in the .ipynb file included in the submission.

2.3 Plots





2.4 Observations

From the plot we can see that the magnitude response is almost maximum at the point $\Theta = 2\pi F_1 T_s$ which matches with our theoretical expectation.

3 Question 2

Excite the above resonator (“filter”) with a periodic source excitation of $F_0 = 160$ Hz. You can approximate the source signal by a narrow-triangular pulse train. Compute the output of the source-filter system over the duration of 0.5 second using the difference equation implementation of the LTI system. Plot the time domain waveform over a few pitch periods so that you can observe waveform characteristics. Play out the 0.5 sec duration sound and comment on the sound quality.

3.1 Solution 2

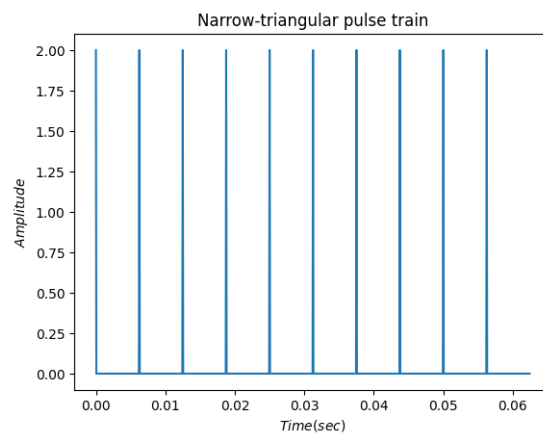
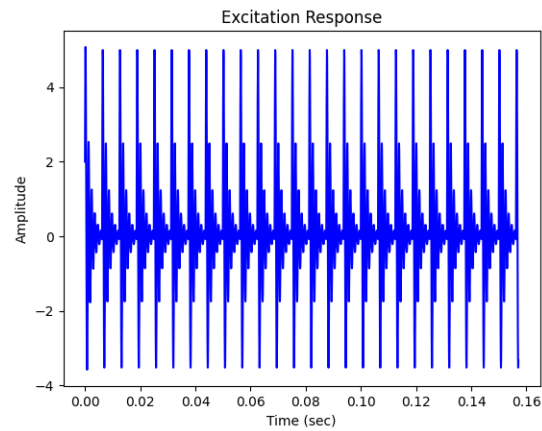
The input waveform was approximated using a square-wave having duty-cycle=0.01.

Rest the calculation is same as in Question 1.

3.2 Code

The code files are included in the .ipynb file included in the submission.

3.3 Plots



3.4 Observations

The sound is very noisy and difficult to predict but seems to resemble vowel a.

4 Question 3

Vary the parameters as indicated below; plot and comment on the differences in waveform and in sound quality for the different parameter combinations.

- (a) $F_0 = 120$ Hz, $F_1 = 300$ Hz, $B_1 = 120$ Hz
- (b) $F_0 = 120$ Hz, $F_1 = 1100$ Hz, $B_1 = 200$ Hz

(c) $F_0 = 180$ Hz, $F_1 = 300$ Hz, $B_1 = 120$ Hz

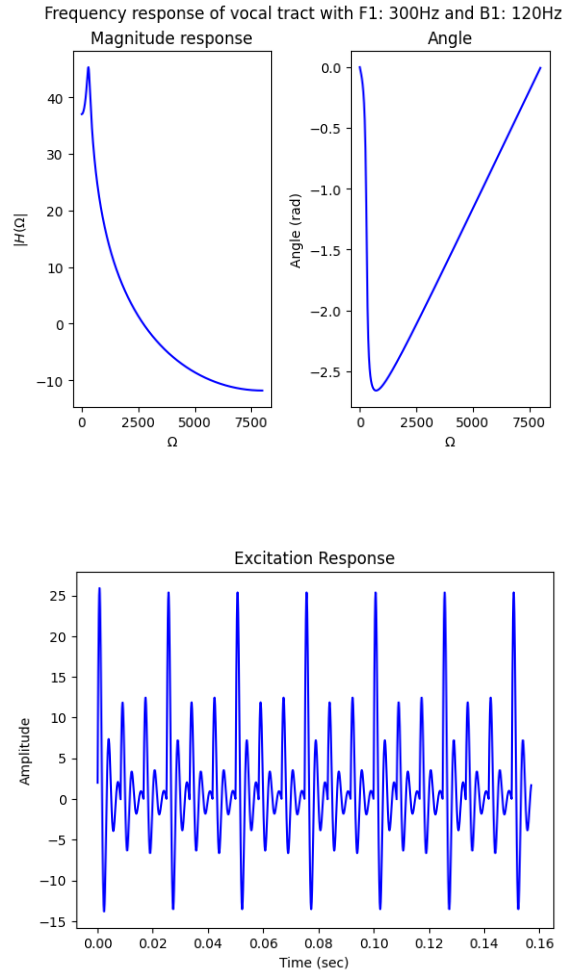
4.1 Solution 3

Same code as the previous 2 parts just convert it to a function and keep changing the variables as given in the question.

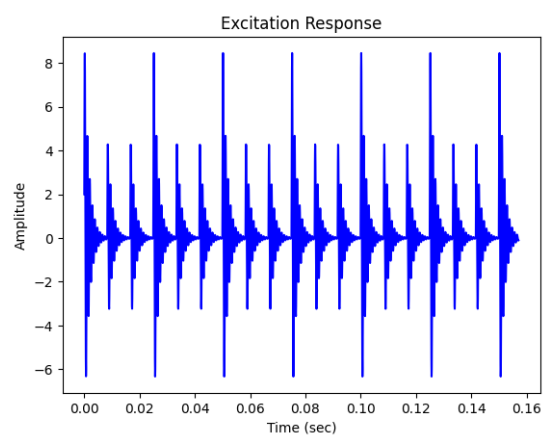
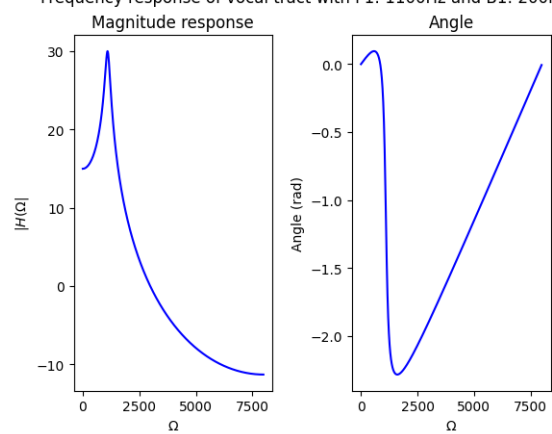
4.2 Code

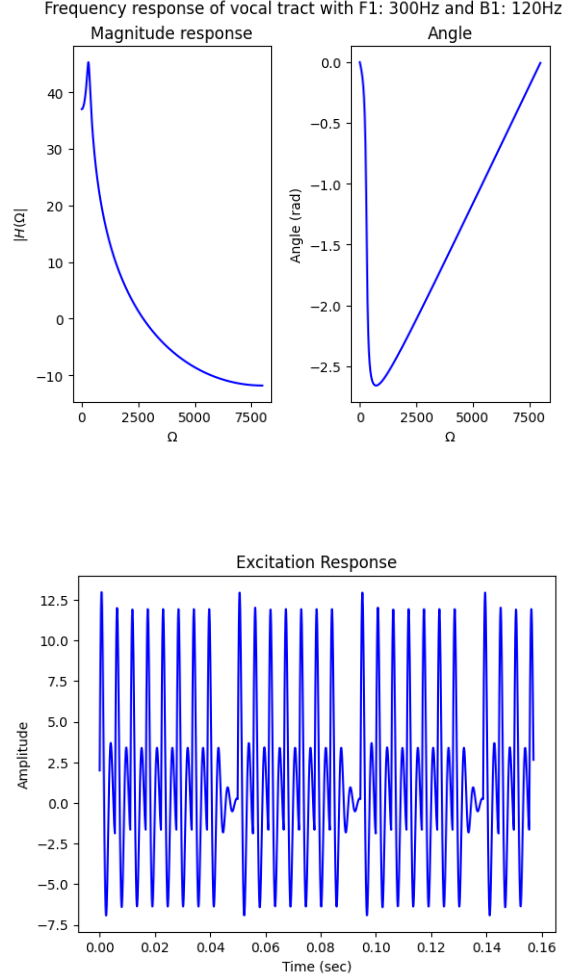
The code files are included in the .ipynb file included in the submission.

4.3 Plots



Frequency response of vocal tract with F1: 1100Hz and B1: 200Hz





4.4 Observations

Although the formant frequencies (and consequently the filter, i.e. the vocal tract characteristics) remain the same in the scenario where the excitation signal frequency (F_0) changes from 120 Hz to 180 Hz, the pitch is higher in the scenario where $F_0 = 180$ Hz because the source excitation—the glottal vibration—controls the perceived pitch.

The vocal tract's features alter when the formant frequency shifts. Thus, even if the frequency of the excitation signal (glottal vibration) is the same, the output waveform seems to be slightly different, with the second waveform's pitch predominating over the first one. In contrast to the second waveform, which has steep dips to the pulse locally, the first waveform has a significantly smoother

time response.

Only one peak in each graph of magnitude response because we are seeing only 1 formant.

5 Question 4

In place of the simple single-resonance signal, synthesize the following more realistic vowel sounds at two distinct pitches ($F_0 = 120$ Hz, $F_0 = 220$ Hz). Keep the bandwidths constant at 100 Hz for all formants. Duration of sound: 0.5 sec. Comment on the sound quality across the different sounds. Plot a few periods of any 2 examples.

Vowel F1, F2, F3

/a/ 730, 1090, 2440

/i/ 270, 2290, 3010

/u/ 300, 870, 2240

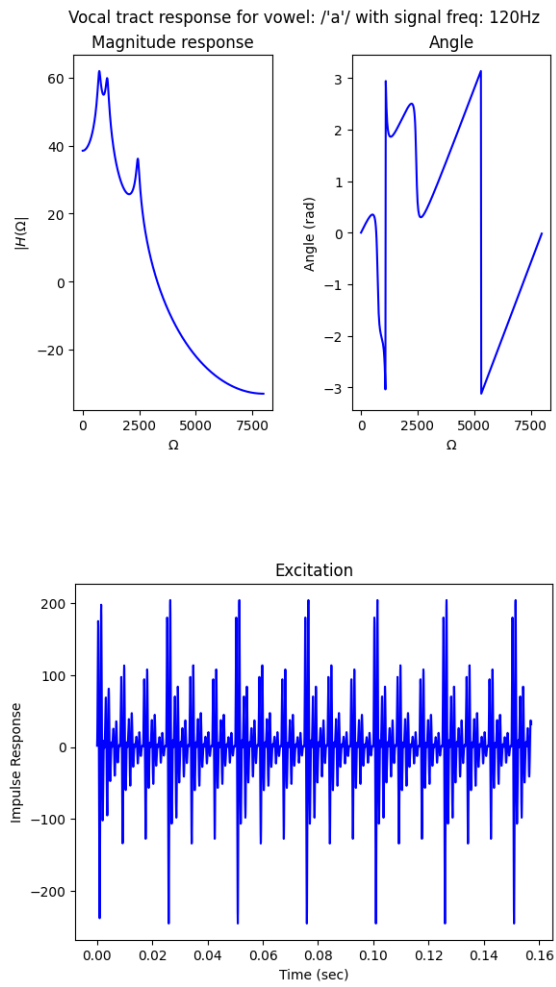
5.1 Solution 4

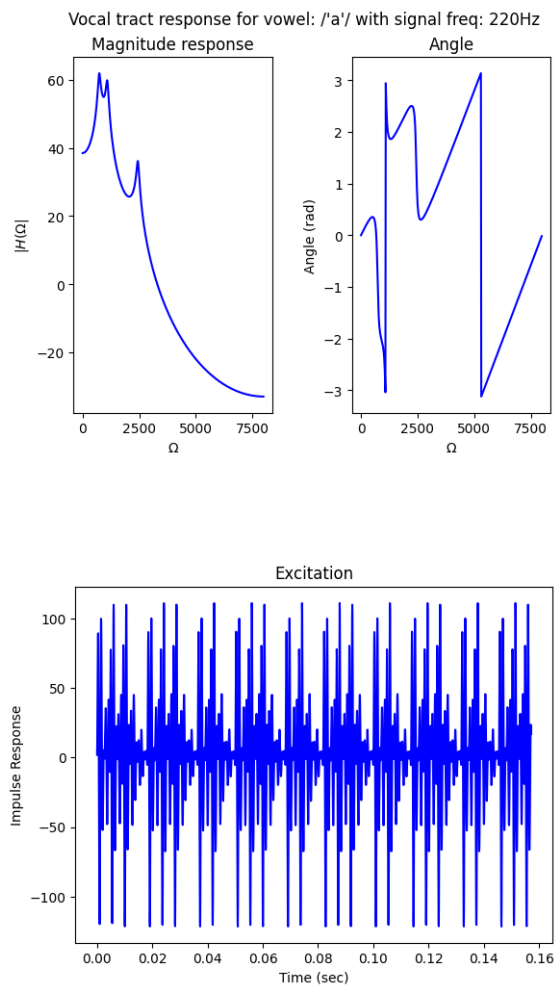
Again the code and method remains same just change the parameters and observe. This time instead of changing the Signal frequency we change the formant frequency for each corresponding signal frequency.

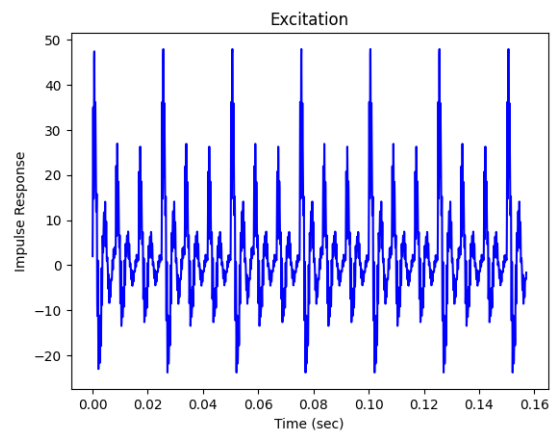
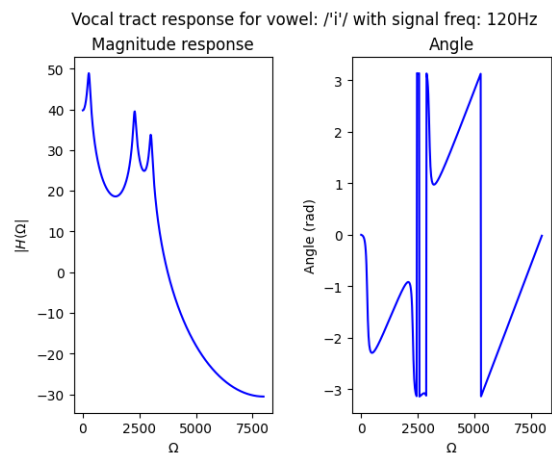
5.2 Code

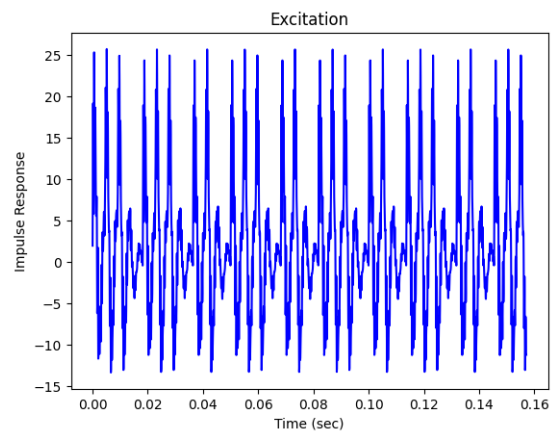
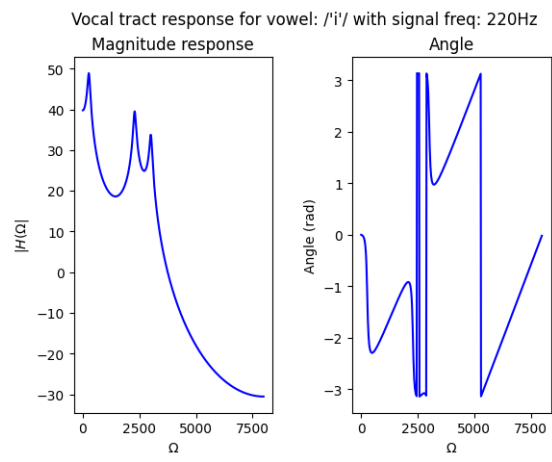
The code files are included in the .ipynb file included in the submission.

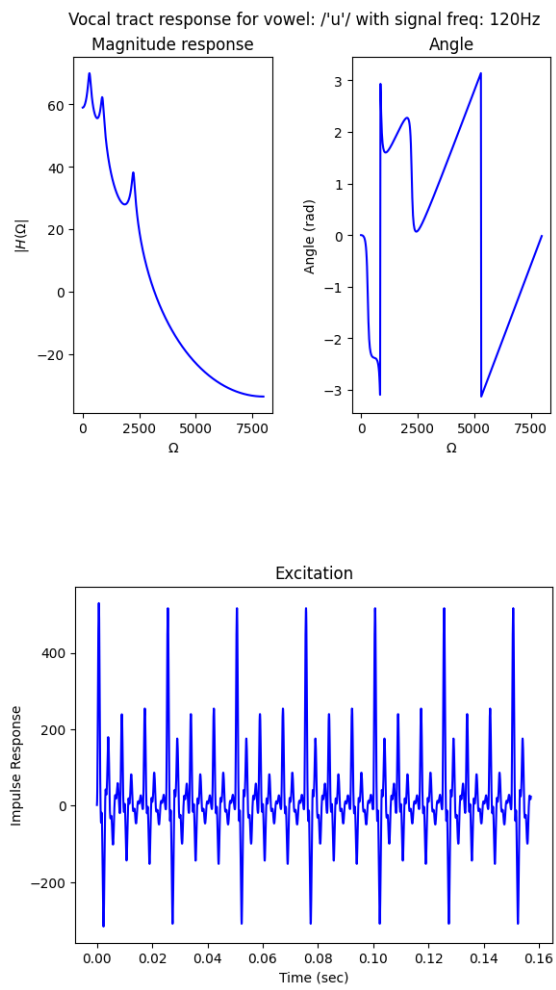
5.3 Plots

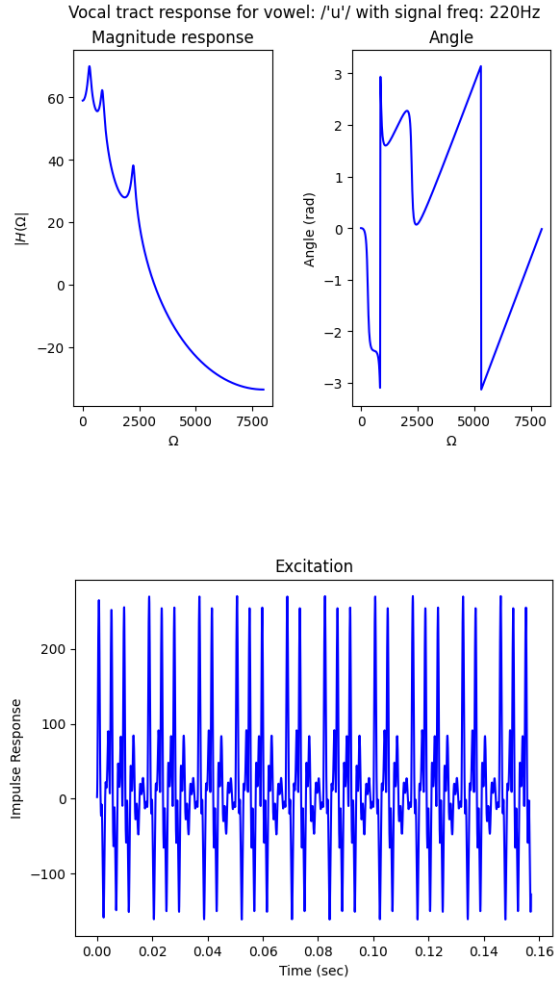












5.4 Observations

Although the vowels can be heard clearly, the noise level remains high. When the F_0 frequencies for the various vowels are changed from 120Hz to 220Hz, the pitch of the sounds changes. According to theory, the voice producing the same vowel at $F_0 = 120\text{Hz}$ should sound like a male voice, whereas the voice producing the same vowel at $F_0 = 220\text{Hz}$ should sound like a female voice. Upon thorough examination, the waves do indeed convey this (to a certain extent, albeit with a rather mechanical voice).

We can clearly see the formant frequencies in the magnitude response as they corresponds to the 3 peaks in each plot.