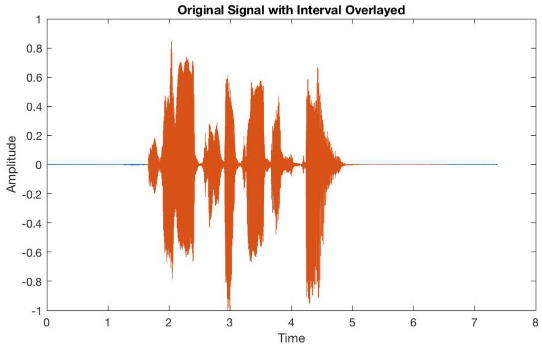
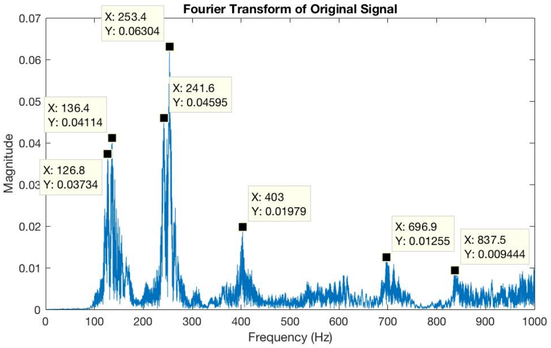
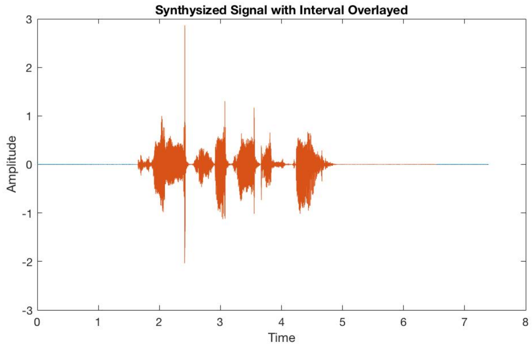
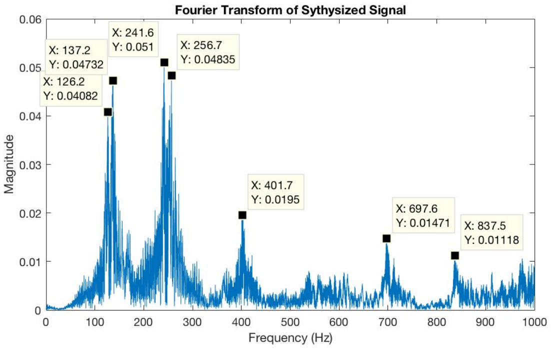
Question (1): use your Fourier Transform function in previous assignments to show the spectra of the original audio signal, and the synthesized audio signal. And comment on you observations on the two spectra. Write your answer between line 21- 30.

Before attempting to plot the Fourier transform, we wanted to gain some intuition into what the spectrum plots should look like. After listening to the recording, it sounds like Professor Ning's voice in both recording versions are pretty normal (regular human range) with the 8000Hz sampling rate entered in the sound() function, however, there are clearly some distortions in the synthesized version. With a little bit of research we found that for males, a typically voice fell between 85 - 160 Hz while talking. Professor Ning's voice, by ear sounds pretty average; not especially deep, not especially high so this range would seem reasonable for the fundamental frequency component of his voice. This means we should expect peaks in this range when looking at the Fourier transforms magnitude. Since the synthesized model seemed to have done a reasonable job at recreating the original recording we should still expect peaks in this range. Looking into how our bodies produce sound, we can also expect harmonic frequencies due to vocal fold vibrations (when they vibrate 2x as fast for the second harmonic, 3x as fast for the third harmonic, etc). Due to this, the spectrum would also contain spikes at 2x and 3x the fundamental frequencies being produced (and so forth for the following harmonics). To view these harmonics we plotted from 0 to 1000Hz to get a good perspective on what is going on (this is the highest note a soprano sings for reference). Now let's see what the spectrum looks like for the entire duration Professor Ning was speaking for.



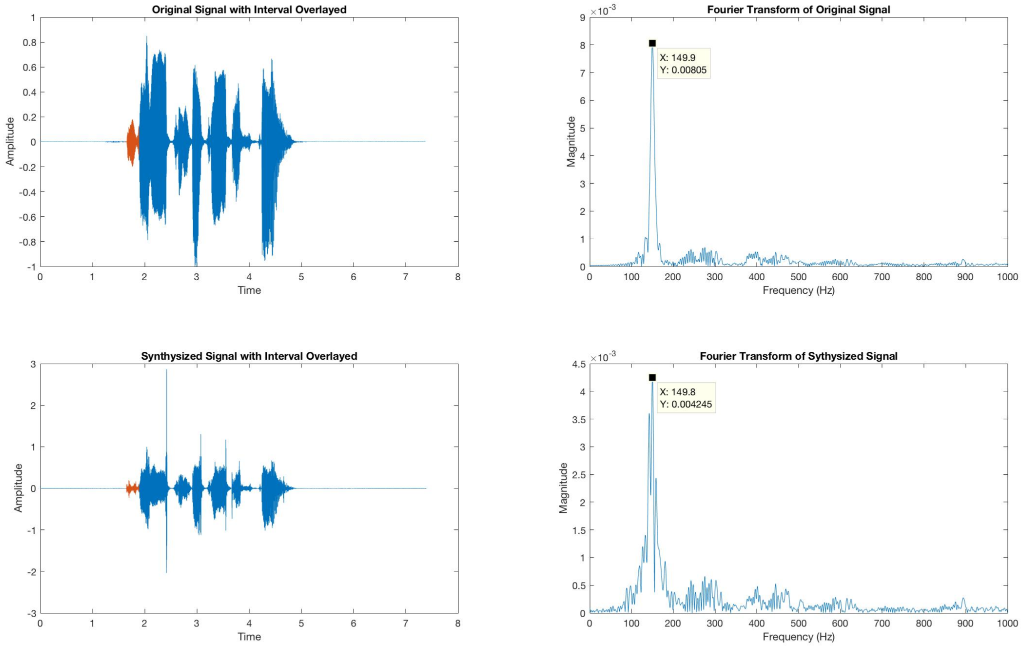




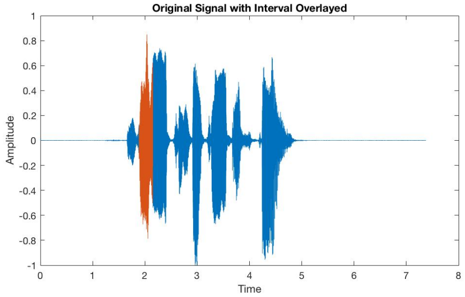


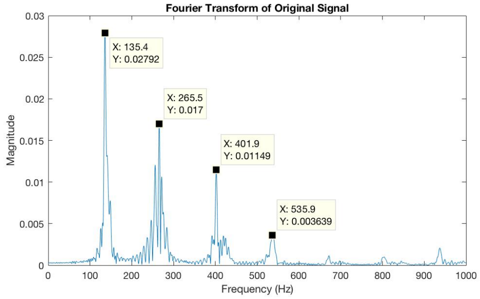
As seen in the above plot, there are big spikes (mainly one big peak centred around the 125-140Hz range) in the 85-160Hz range as we expected. Since an entire sentence is said, there is lots of variation in the frequencies being outputted. We also see the second harmonic frequencies (at 2x the fundamental frequency range) are extremely predominate. All these observations are valid for both the original and synthesized (sorry for spelling it wrong in the plot) but are there any differences between them? Overall it seems like the model did a pretty good job at synthesizing the signal, however, we can see that the second harmonic frequencies are smaller in magnitude and the fundamental frequencies are larger in the synthesized signal. Also, the synthesized signals spectrum is a less smooth plot; this is especially visible in regions like the 180-240Hz range where the variation in magnitude is much higher than in the original signal (even after compensating for the difference in the y axis scale). At a high level this makes sense that the original signal would have a cleaner range plot with smoother flow of frequencies than the synthesized version. Based on the idea that as sound travels from its source at vocal folds through the voice tract and can become louder/softer/change-pitch, it makes sense that this plot looks pretty messy with a bunch spikes in our male fundamental frequency range and the higher harmonics since we are looking at the entire sentence that was spoken and all the temporal information (when which frequencies happened) was lost. To help verify our results we decided to split the sentence up into syllable chunks to isolate vowel sounds and see the spectrum components that are combining to give us the plot shown above and give us an idea of what happened when. The following few pages contain these plots for the following sections of the phrase: ‘B’, ‘M’, ‘E’, ‘2’, ‘5’, ‘2’, ‘is’, ‘fun’

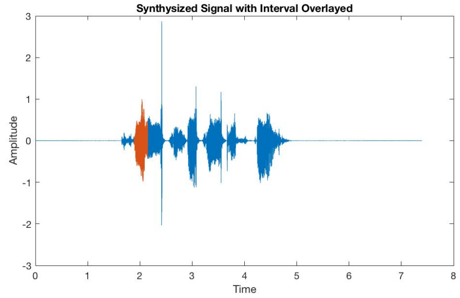
B:

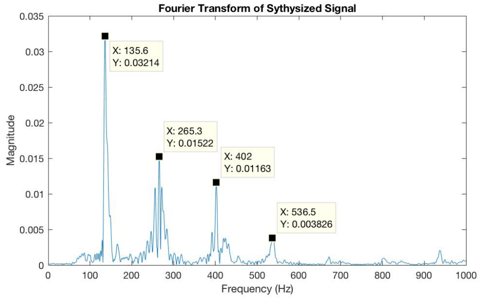


M:

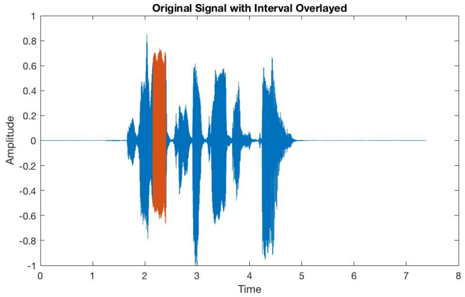

Original Signal with Interval Overlayed

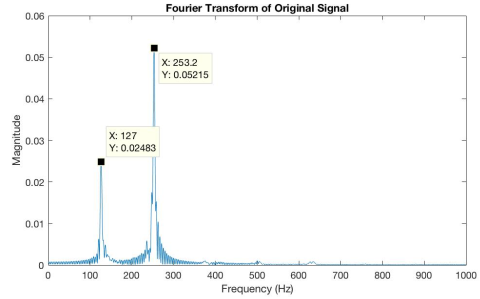

Fourier Transform of Original Signal

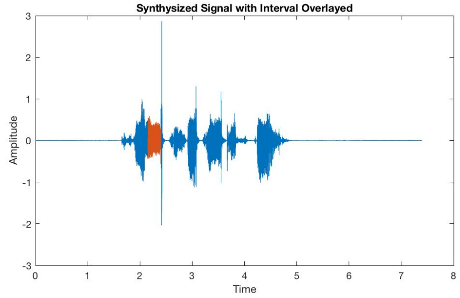

Synthysized Signal with Interval Overlayed

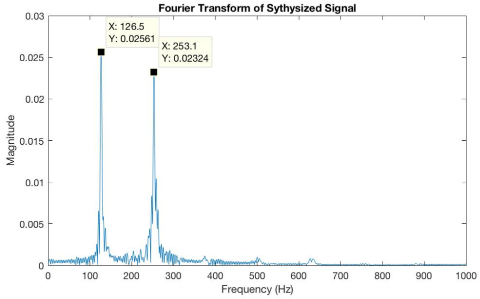

Fourier Transform of Sythysized Signal

E:

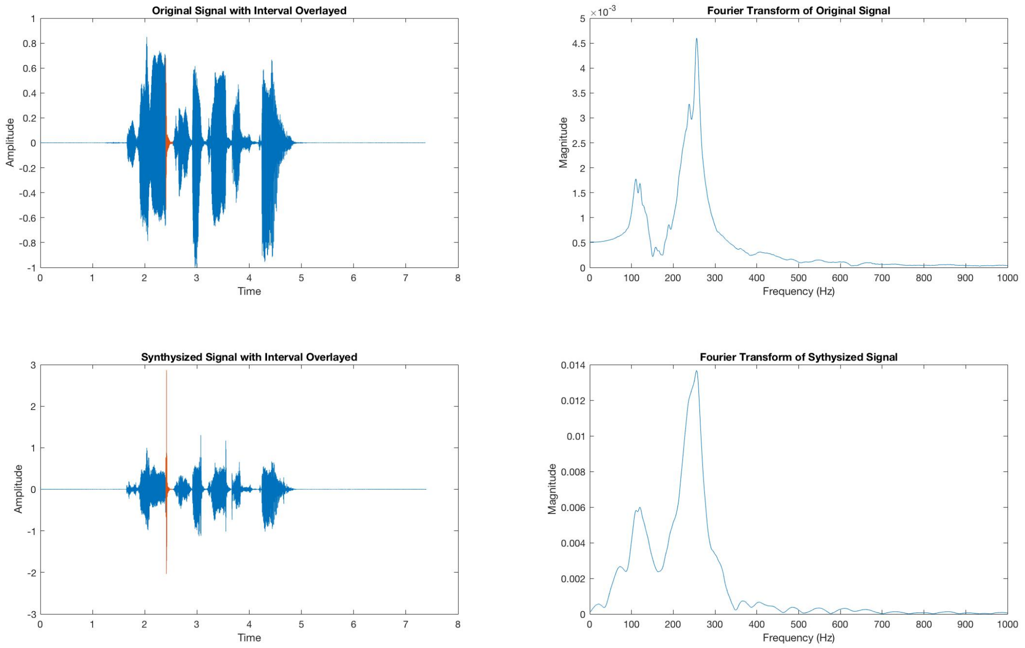

Original Signal with Interval Overlayed


Fourier Transform of Original Signal

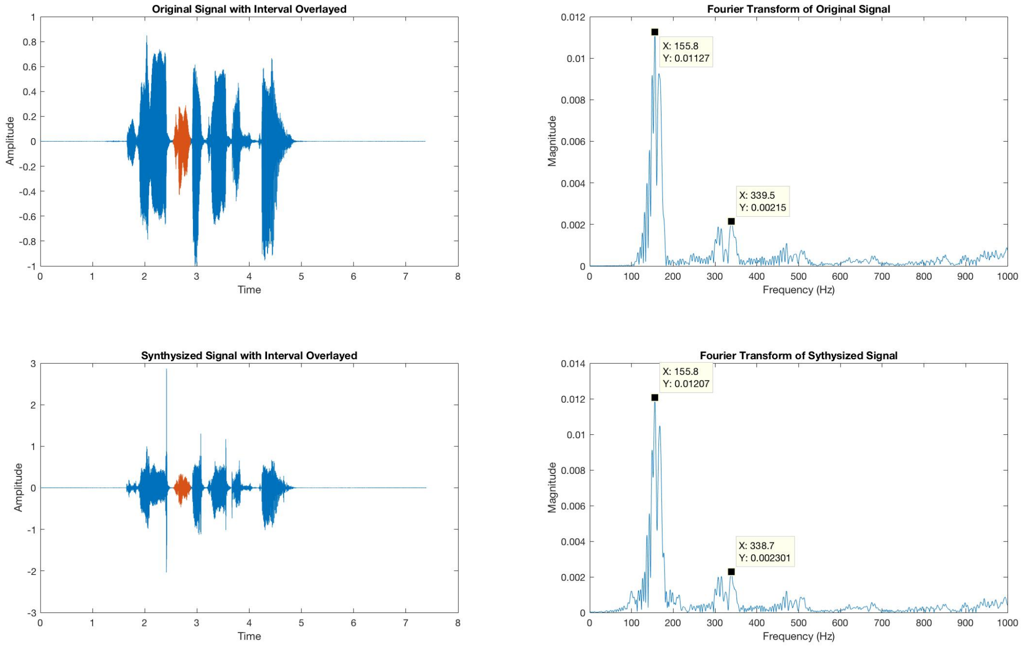

Synthysized Signal with Interval Overlayed


Fourier Transform of Sythysized Signal

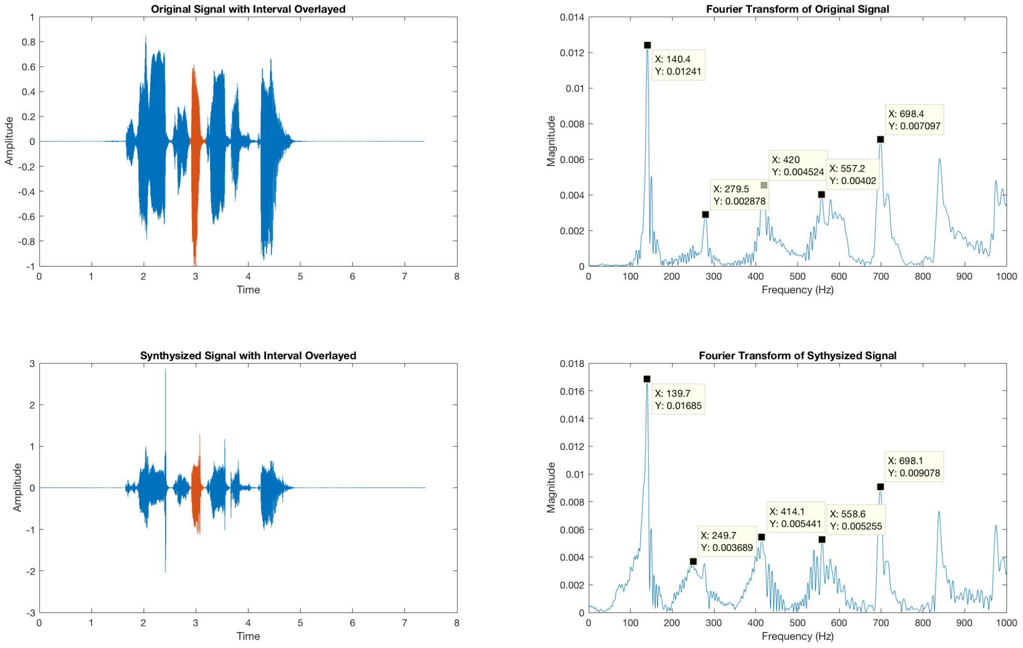
Gap in speech (bonus: does not contain significant magnitude frequency components):



2:

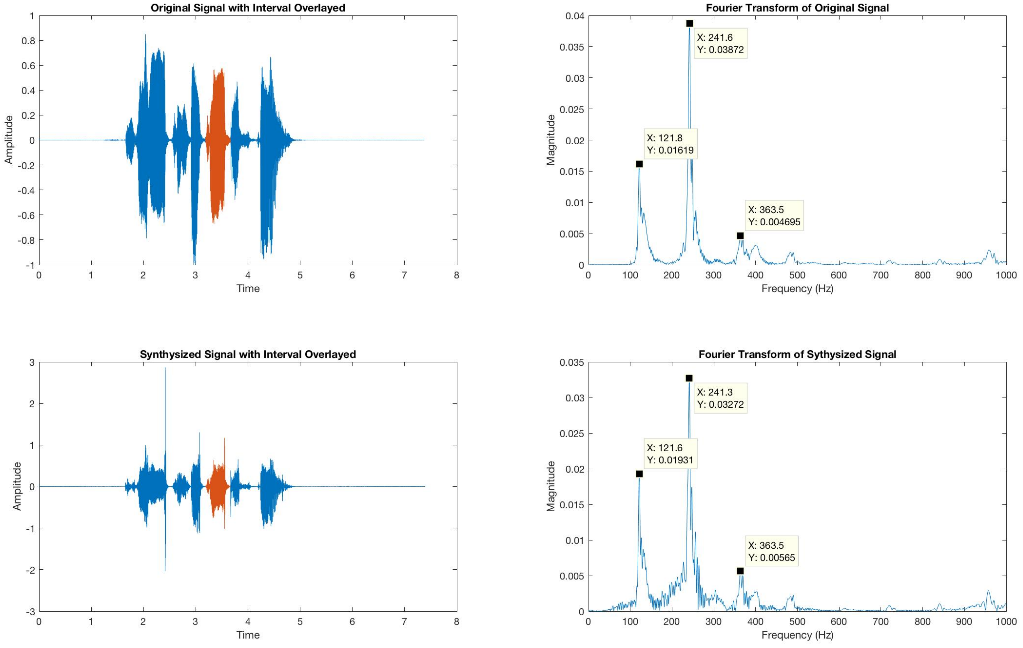


5:



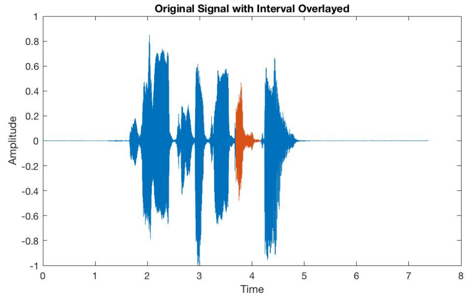
Original Signal with Interval Overlayed
Amplitude
Time
Fourier Transform of Original Signal
Magnitude
Frequency (Hz)
X: 140.4
Y: 0.01241
X: 279.5
Y: 0.002878
X: 420
Y: 0.004524
X: 557.2
Y: 0.00402
X: 698.4
Y: 0.007097
Synthysized Signal with Interval Overlayed
Amplitude
Time
Fourier Transform of Sythysized Signal
Magnitude
Frequency (Hz)
X: 139.7
Y: 0.01685
X: 249.7
Y: 0.003689
X: 414.1
Y: 0.005441
X: 558.6
Y: 0.005255
X: 698.1
Y: 0.009078

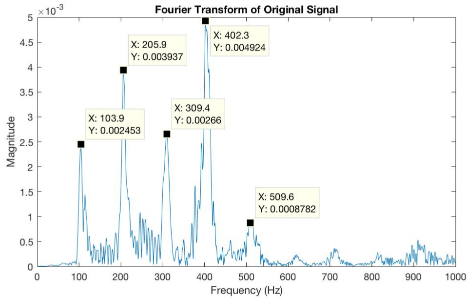
2:

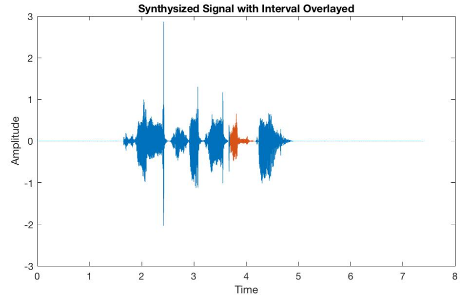


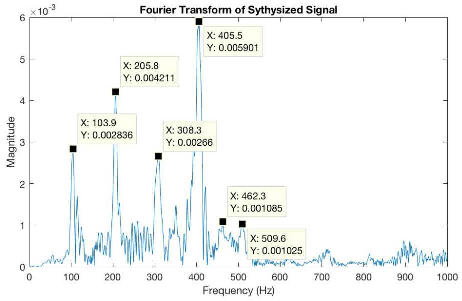
Original Signal with Interval Overlayed
Amplitude
Time
Fourier Transform of Original Signal
Magnitude
Frequency (Hz)
X: 241.6
Y: 0.03872
X: 121.8
Y: 0.01619
X: 363.5
Y: 0.004695
Synthysized Signal with Interval Overlayed
Amplitude
Time
Fourier Transform of Sythysized Signal
Magnitude
Frequency (Hz)
X: 241.3
Y: 0.03272
X: 121.6
Y: 0.01931
X: 363.5
Y: 0.00565

Is:

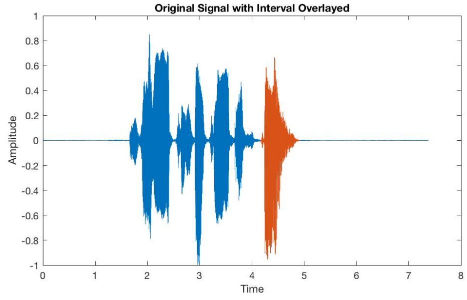

Original Signal with Interval Overlayed
Amplitude
Time
1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1
0 1 2 3 4 5 6 7 8

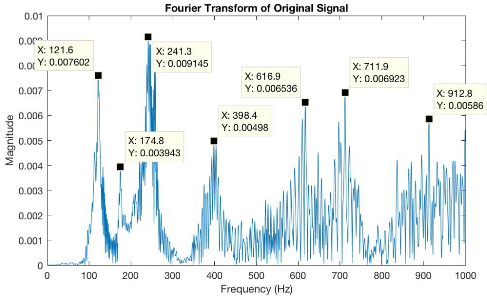

Fourier Transform of Original Signal
×10^-3
Magnitude
Frequency (Hz)
5 4.5 4 3.5 3 2.5 2 1.5 1 0.5 0
0 100 200 300 400 500 600 700 800 900 1000
X: 103.9
Y: 0.002453
X: 205.9
Y: 0.003937
X: 309.4
Y: 0.00266
X: 402.3
Y: 0.004924
X: 509.6
Y: 0.0008782

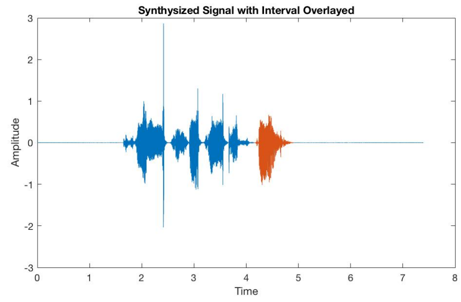

Synthysized Signal with Interval Overlayed
Amplitude
Time
3 2 1 0 -1 -2 -3
0 1 2 3 4 5 6 7 8

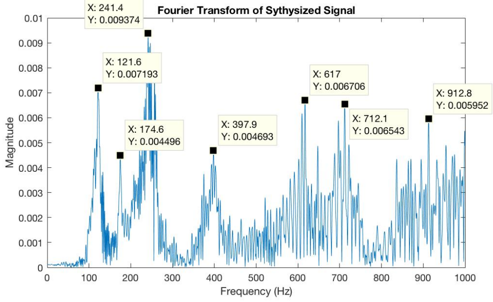

Fourier Transform of Sythysized Signal
×10^-3
Magnitude
Frequency (Hz)
6 5 4 3 2 1 0
0 100 200 300 400 500 600 700 800 900 1000
X: 103.9
Y: 0.002836
X: 205.8
Y: 0.004211
X: 308.3
Y: 0.00266
X: 405.5
Y: 0.005901
X: 462.3
Y: 0.001085
X: 509.6
Y: 0.001025

Fun:


Original Signal with Interval Overlayed
Amplitude
Time
1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1
0 1 2 3 4 5 6 7 8


Fourier Transform of Original Signal
Magnitude
Frequency (Hz)
0.01 0.00 0.00 0.007 0.006 0.005 0.004 0.003 0.002 0.001 0
0 100 200 300 400 500 600 700 800 900 1000
X: 121.6
Y: 0.007602
X: 241.3
Y: 0.009145
X: 174.8
Y: 0.003943
X: 398.4
Y: 0.00498
X: 616.9
Y: 0.006536
X: 711.9
Y: 0.006923
X: 912.8
Y: 0.00586


Synthysized Signal with Interval Overlayed
Amplitude
Time
3 2 1 0 -1 -2 -3
0 1 2 3 4 5 6 7 8


Fourier Transform of Sythysized Signal
Magnitude
Frequency (Hz)
0.01 0.009 0.008 0.007 0.006 0.005 0.004 0.003 0.002 0.001 0
0 100 200 300 400 500 600 700 800 900 1000
X: 241.4
Y: 0.009374
X: 121.6
Y: 0.007193
X: 174.6
Y: 0.004496
X: 397.9
Y: 0.004693
X: 617
Y: 0.006706
X: 712.1
Y: 0.006543
X: 912.8
Y: 0.005952