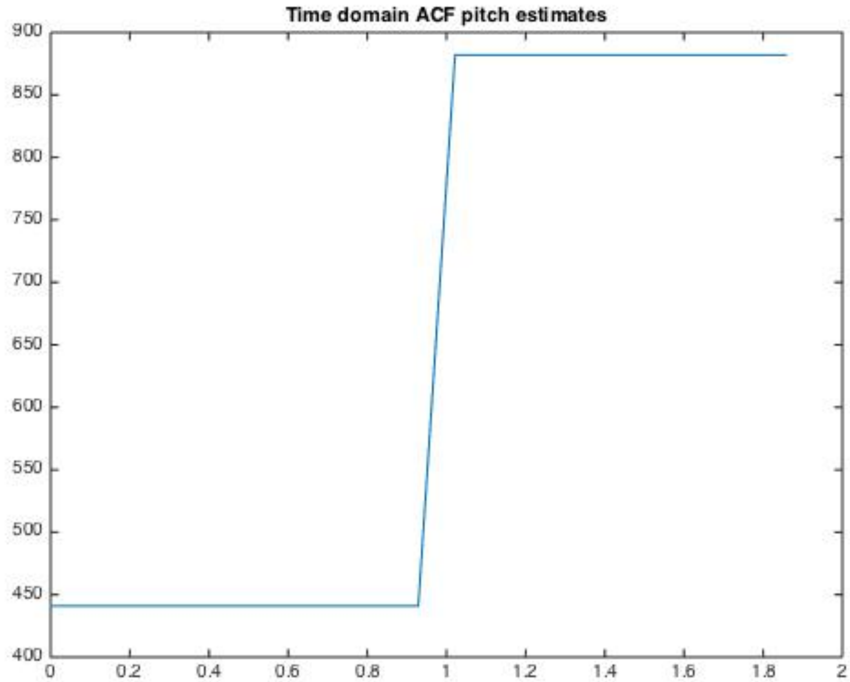
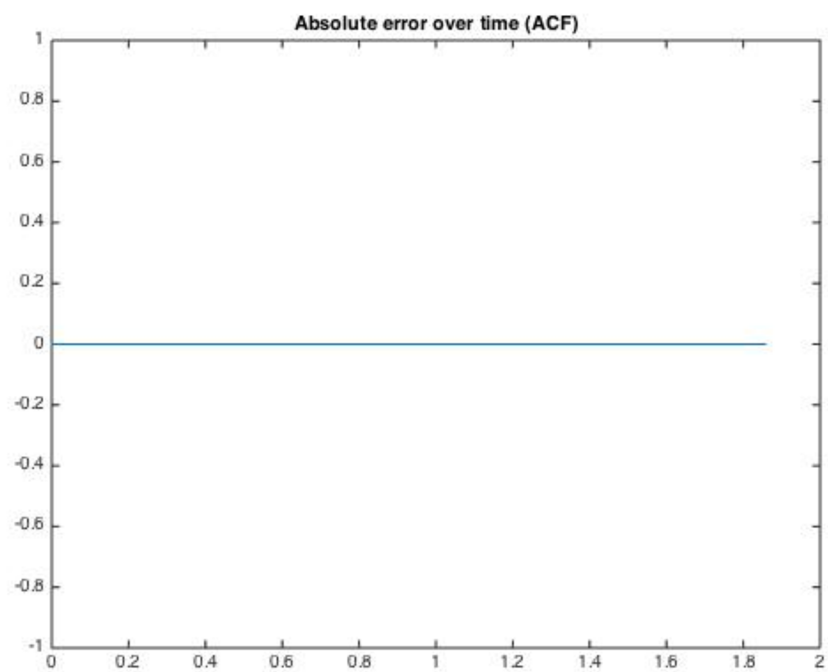
Chris Laguna
Ying Zhan

MUS 6201 Homework 1

1)

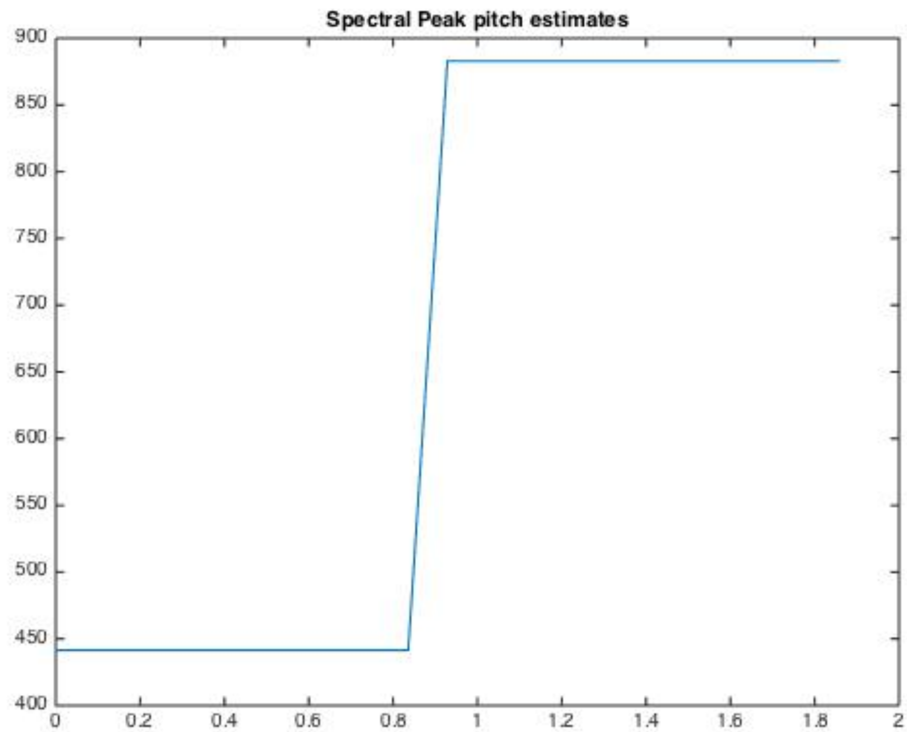


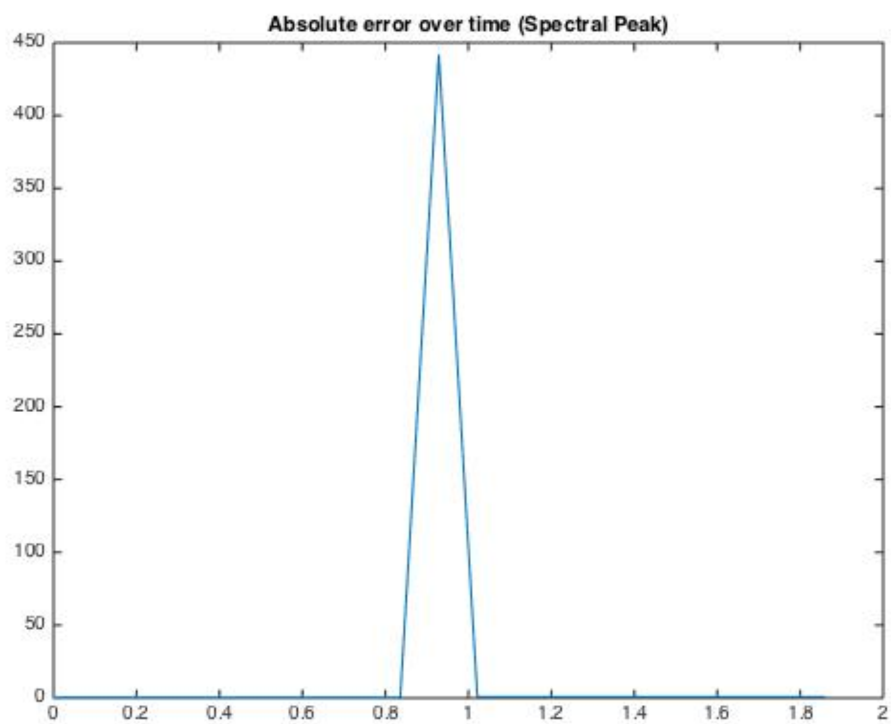


Window size = 8192, Hop size = 4096.
There is no error. This is because the sampling frequency is an integer multiple of the fundamental frequency.

Chris Laguna
Ying Zhan

2)





Chris Laguna
Ying Zhan

Window size = 8192, Hop size = 4096.

We have one block where our error is quite large, at the time when the sinusoidal frequency changes from 441Hz to 882 Hz. This is because that block contains segments of time with different fundamental frequencies (441 and 882). By blocking the signal, we assume that there is only one fundamental frequency within that block: in this case, that is not true, so really the ground truth is undefined. However, it appears that for the majority of the block, the fundamental frequency was 882, but we defined the ground truth to be 441. This is not that important of an error.

You can't really see from the graph, but we do have an error of approximately 0.407 Hz during the 441Hz sinusoidal and 0.861Hz during the 882Hz sinusoidal. This error occurs because our fundamental frequency choices are essentially quantized based on the frequency resolution of our FFT, so we can't get the exact frequency.

Because the audio signal is sinusoidal, the magnitude spectrum will be nonzero only at a single bin (and it's negative-frequency equivalent bin, ignoring spectral leakage), so this method should work for sinusoidals. This shouldn't be the case in general, so we expect this method to fail during the test-set evaluation.

3)

Mean ACF RMS Error: **463.1559** cents.
Mean Spectral Peak RMS Error: **1933.436** cents.

The time domain ACF method gives a very accurate fundamental frequency estimate, sometimes. When it fails, it fails hard, predicting either the maximum or minimum frequencies (we defined a frequency search range). This can be improved by post-processing the estimates with a median filter to remove outliers, smoothing the autocorrelation function over time, and pre-processing the audio signal by band-pass filtering it based on the frequency search range.

The max spectral peak method doesn't ever give an accurate estimate, but it seems to be 'better' at failing. Neither method seems to work very well. Better frequency domain pitch estimation methods exist such as doing an autocorrelation of the magnitude spectrum (you should get a peak corresponding to the fundamental frequency if there are integer harmonics, as is true in most pitched sounds). The improvements that we suggested above for the time domain ACF method also apply to frequency domain ACF based pitch estimation.

We were able to get the time-domain error much lower than the frequency-domain error by setting the search range to 100-1000Hz based on the frequencies that exist in the test-set. This is an example of overfitting…and won't work in general.