

Zach Kondak and Ryan Rose

MUSI 6201

10 October 2017

Assignment 3 - Monophonic Pitch Tracking Revisited

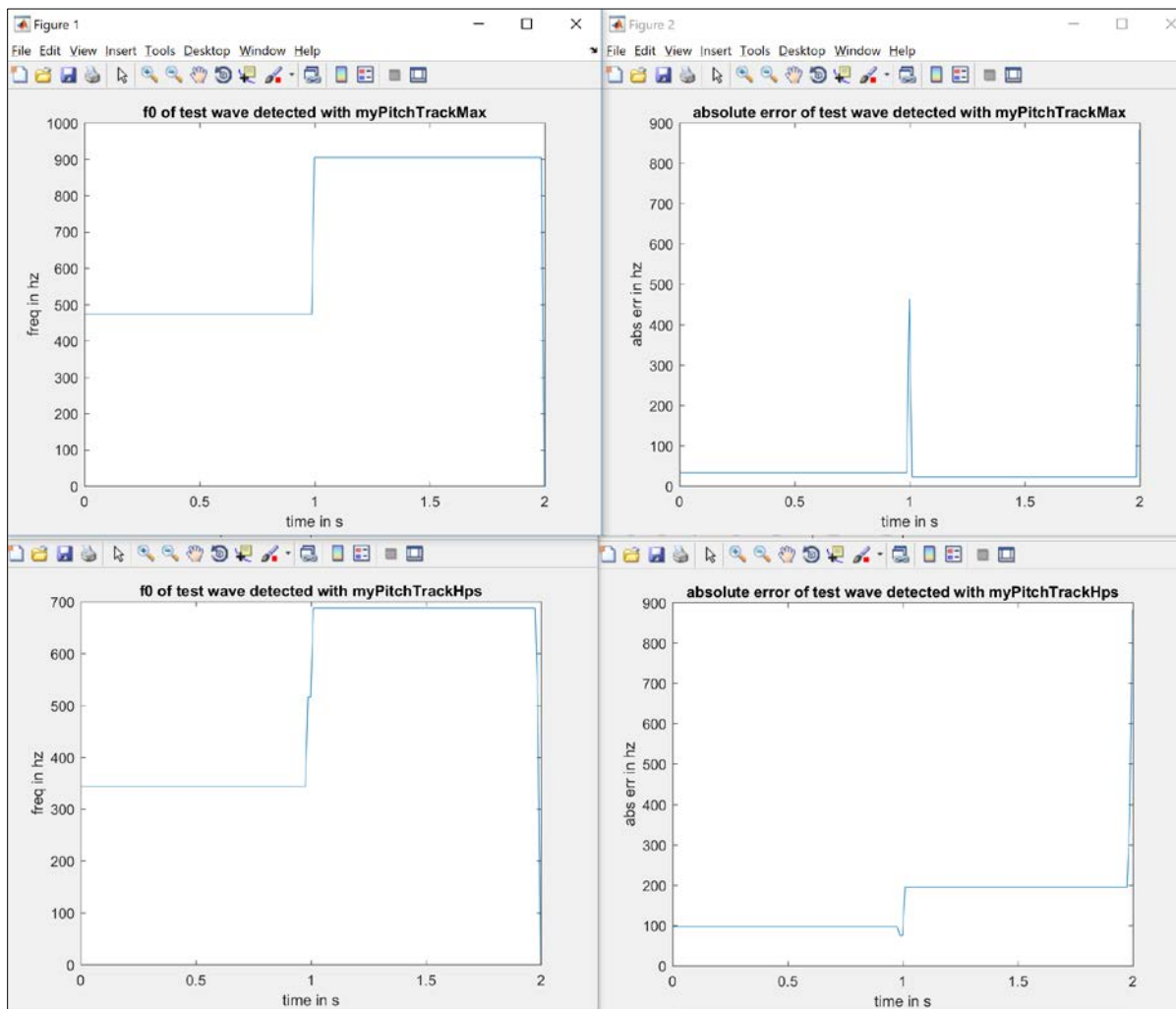
A. Maximum spectral peak based pitch tracker

If the blockSize = 1024 in myPitchTrackMax(), what is the resolution of your pitch tracker? Can this be improved without changing the block-size? If yes, how? If no, why?

Since blockSize is being used as the FFT length, which will determine resolution, the resolution is equal to $\text{blockSize} / 1024$. Yes: if the FFT length is increased, the resolution will be increased, at the expense of some error introduced by zero-padding or interpolation of the block undergoing FFT.

D. Evaluation

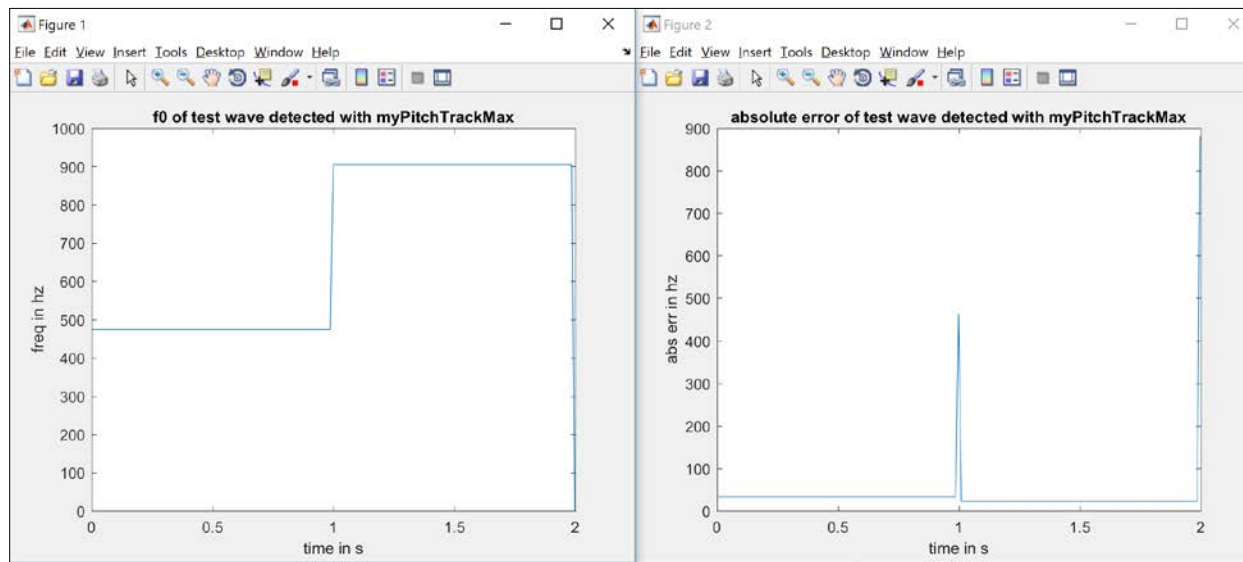
Plot the absolute error per block and discuss the possible causes for the deviation. Repeat for myPitchTrackHPS() with the same signal and parameters. Why does the HPS method fail with this signal?



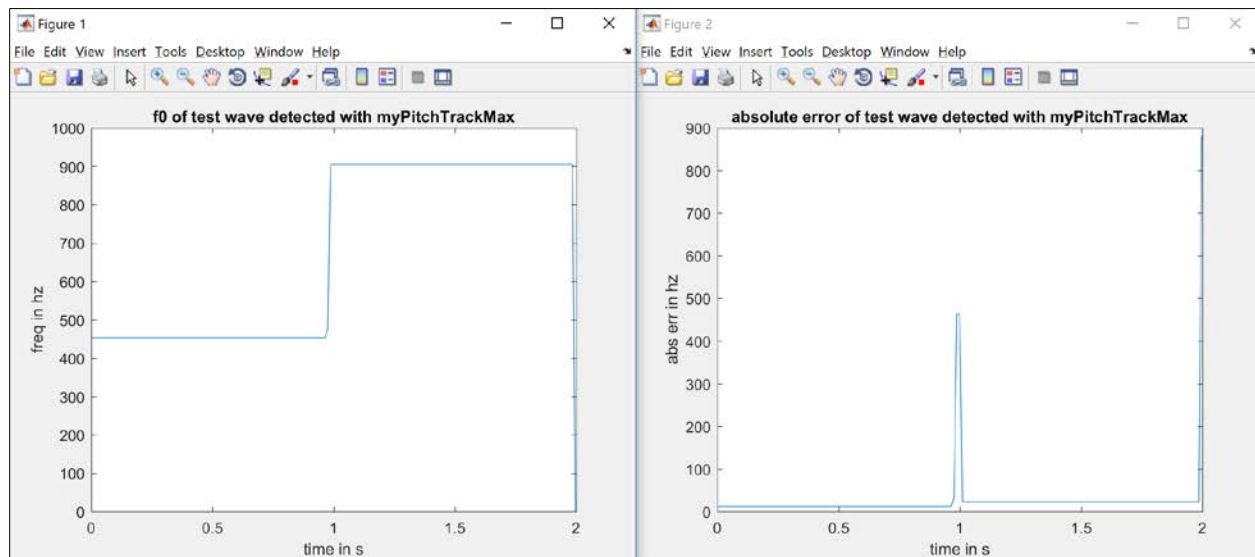
Similar to the ACF method of pitch tracking, myPitchTrackMax() has trouble with instantaneous frequency change and thus misidentified the pitch when it changed, resulting in a large error spike equal to the change of frequency. For myPitchTrackHps, in general, the error is quite high, though it was able to handle the frequency change without spiking. HPS fails with this signal because it's meant for signals accompanied by their associated harmonics – since there are no harmonics with this sine wave, the multiplication with missing harmonics will cause this method to fail.

Next use (blockSize = 2048, hopSize = 512) and repeat the above experiment (only for the max spectra method). Do you see any improvement in performance?

blockSize = 1024



blockSize = 2048



There is some improvement when the larger block size is used. There is less error in predicting f0 when the pitch is in steady-state, however, the larger block size created a wider spike in the transition period.

Evaluate your `myPitchTrackMax()` using the training set (see attachment) and the `myEvaluation2()` method (use `blockSize = 1024`, `hopSize = 512`). Report the average performance metrics across the training set.

All values are 1.0e+03 * ...	<i>myPitchTrackMax</i>			<i>myPitchTrackHps</i>		
	rms	pfp	pfn	rms	pfp	pfn
<i>01-D_AMairena</i>	1.7707	0.1000	0.0001	1.6160	0.1000	0.0001
<i>24-M1_AMairena-Martinete</i>	2.0473	0.0992	0	2.0237	0.0992	0
<i>63-M2_AMairena</i>	2.1246	0.0994	0	1.6633	0.0994	0

E. Voicing Detection

Evaluate your `myPitchTrack()` using the training set (see attachment) and the `myEvaluation2()` method (use `blockSize = 1024`, `hopSize = 512`) over all 3 pitch trackers (acf, max and hps) and report the results with two values of threshold (threshold = -40, -20).

All values are 1.0e+03 * ...	<i>myPitchTrack - ACF (-40 threshold)</i>			<i>myPitchTrack - ACF (-20 threshold)</i>		
	rms	pfp	pfn	rms	pfp	pfn
<i>01-D_AMairena</i>	4.9782	0	0.0590	6.5726	0	0.1000
<i>24-M1_AMairena-Martinete</i>	5.3534	0	0.0747	6.2312	0	0.1000
<i>63-M2_AMairena</i>	3.9942	0	0.0381	6.5677	0	0.1000

All values are 1.0e+03 * ...	<i>myPitchTrack - MAX (-40 threshold)</i>			<i>myPitchTrack - MAX (-20 threshold)</i>		
	rms	pfp	pfn	rms	pfp	pfn
<i>01-D_AMairena</i>	5.0840	0	0.0590	6.5726	0	0.1000
<i>24-M1_AMairena-Martinete</i>	5.4106	0	0.0747	6.2312	0	0.1000
<i>63-M2_AMairena</i>	4.3377	0	0.0381	6.5677	0	0.1000

All values are 1.0e+03 * ...	<i>myPitchTrack - HPS (-40 threshold)</i>			<i>myPitchTrack - HPS (-20 threshold)</i>		
	rms	pfp	pfn	rms	pfp	pfn
<i>01-D_AMairena</i>	5.0911	0	0.0590	6.5726	0	0.1000
<i>24-M1_AMairena-Martinete</i>	5.4365	0	0.0747	6.2312	0	0.1000
<i>63-M2_AMairena</i>	4.2000	0	0.0381	6.5677	0	0.1000