Assignment 1: Pitch Tracking

Amruta Vidwans and Rithesh Kumar

1.Blockwise autocorrelation based pitch tracker

Generate a test signal (sine, f = 441 Hz from 0-1 sec and f = 882 Hz from 1-2 sec), apply the pitch tracker (expected error <= 5%) and plot the f0 curve. Plot the absolute error and discuss the possible causes for the deviation.

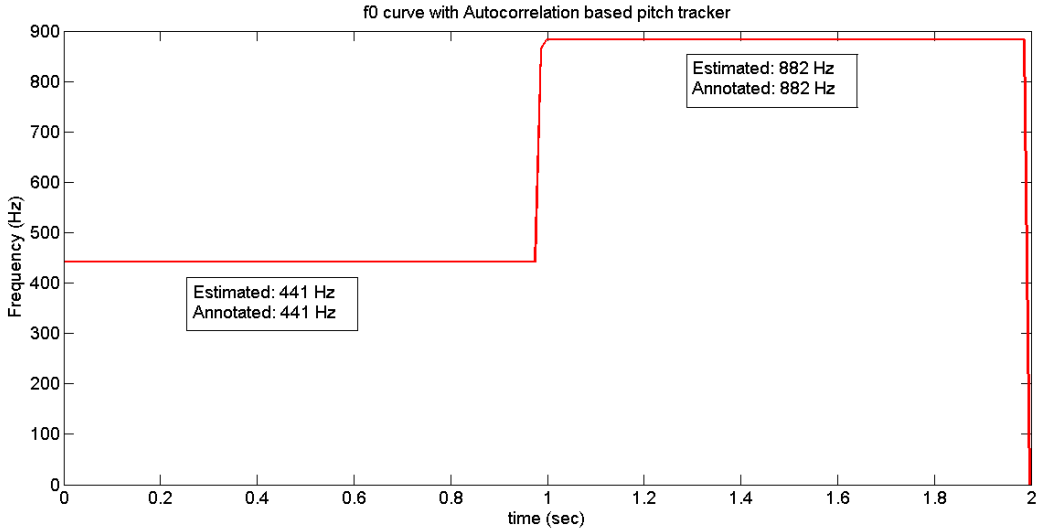

Figure 3. Autocorrelation based pitch tracker f0 curve with window size 1024 and hop size 512 samplesand fs as 44100Hz

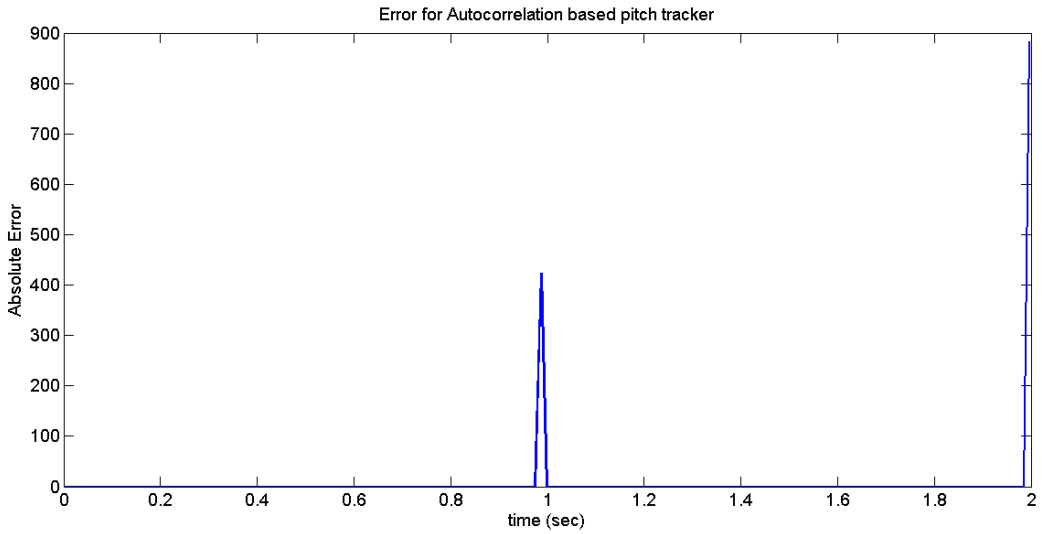

Figure 4. Autocorrelation based pitch tracker absolute error with respect to annotationwith window size 1024, hop size 512 samples and fs as 44100Hz

This pitch tracker as seen from figure 3 and figure 4 has error of 0%. Experiments were also tried by changing the window size and hop length to give similar result. Window size should be atleast equal to one period of the signal for accurate estimation of the pitch.
The 1st sudden peak in the figure 4 is at the change point in the signal from frequency 441Hz to 882Hz, since the window will have both 441 Hz and 882 Hz present in it and the strongest peak will be selected. This will depend on which frequency is occupying the window more. The peak in the end corresponds to error at 882Hz as there was not enough signal in the window to estimate the pitch.
Autocorrelation based method gives better results as we calculate the pitch by considering the lag time where the signal overlaps with itself the most (2nd peak in the autocorrelation). This step is limited by the shift with which we overlap the signal and calculate the area, which here is just a single sample shift. This makes the pitch estimation same as the actual and hence more accurate.

2. [40] Blockwise maximum spectral peak based pitch tracker
Use the same test signal above with the same parameters (ex. windowSize, hopSize, fs… please specify them in your report), apply the pitch tracker (expected error <= 5%) and plot the f0 curve. Plot the absolute error and discuss the possible causes for the deviation.

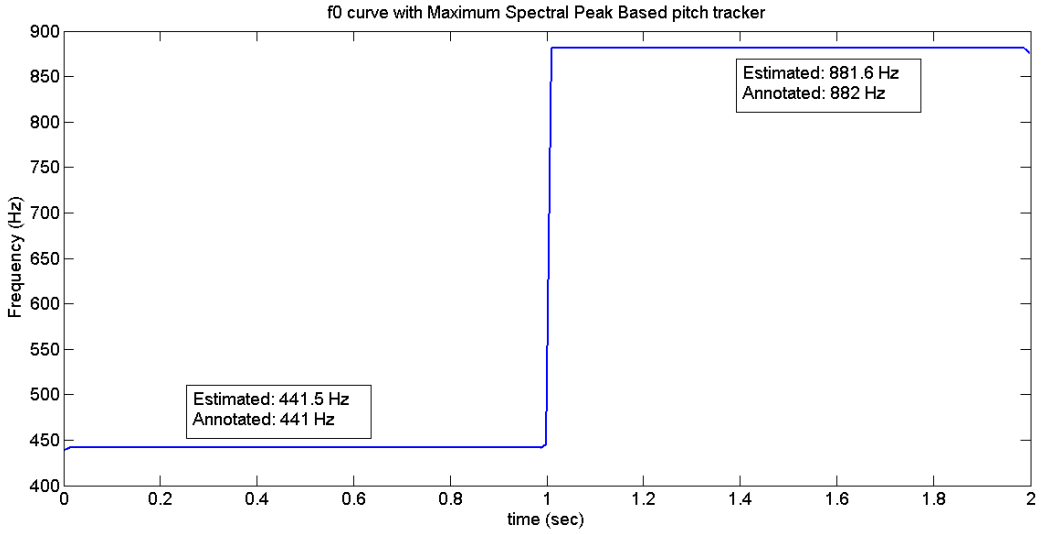

Figure 1. Maximum Spectral Peak based pitch tracker f0 curve with window size 1024 and hop size 512 samplesand fs as 44100Hz

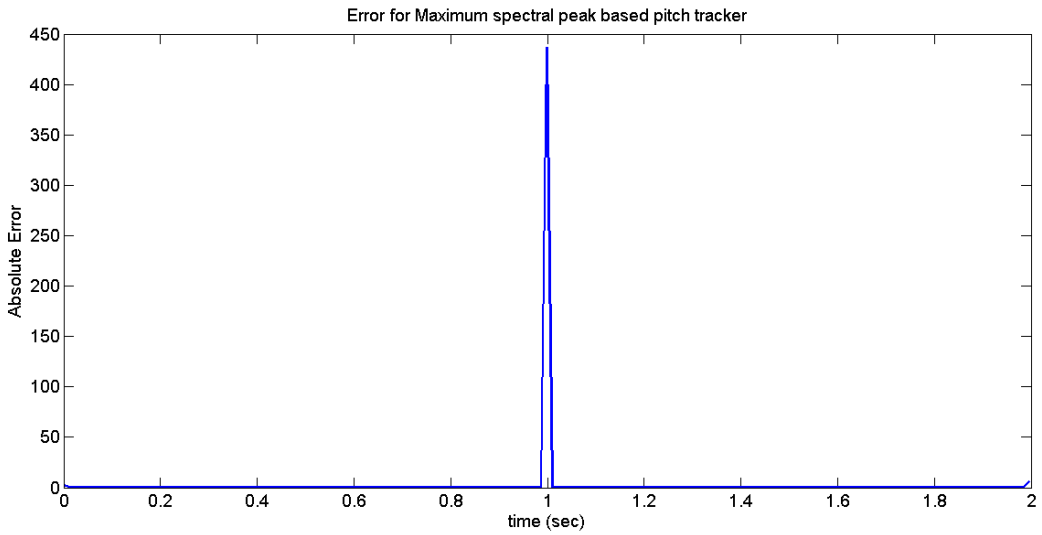

Figure 2. Maximum Spectral Peak based pitch tracker absolute error with respect to annotationwith window size 1024, hop size 512 samples and fs as 44100Hz

This pitchtracker as seen from figure 1 and figure 2 has error of <5% where the 441Hz and 882Hz frequencies are present constant for 1 sec. Experiments were also tried by changing the window size and hop length to give similar result. The fft computation is done by taking transform length 32 times the window size as this increases the frequency resolution and the error was ~0.1% as seen in figure 1. The fft computation when done by taking the transform length to be same as window size, it gives ~2% error. The window size is 1024 and the fft transform length when kept same as window size, the frequency resolution will be fs/window size and hence will have quantized resolution with each step of 43.15Hz, thus limiting the accuracy.
The sudden peak in the figure 2 is at the change point in the signal from frequency 441Hz to 882Hz, since the window will have both 441 Hz and 882 Hz present in it and the strongest peak will be selected. This will depend on which frequency is occupying the window more.

3. [20] Evaluation (windowSize = 1024, hopSize = 512)
Evaluate the above methods (ACF, MaxSpec) using the training set (see attachment). Please report the overall errCent_rms of the training set for each method. What are the differences between the two methods? What are the potential solutions to improve their performances?

The RMS error for each of song for both the methods is as below: (where errCent_rms_ACF is error of autocorrelation based method and errCent_rms_MaxSpec is error of maximum spectral peak based method)
63-M2_AMairena
errCent_rms_ACF 2.256888412450436e+03
errCent_rms_MaxSpec 2.144285072986754e+03

24-M1_AMairena-Martinete
errCent_rms_ACF 2.584366254660974e+03
errCent_rms_MaxSpec 2.075228347582959e+03

01-D_AMairena

errCent_rms_ACF 1.910e+03
errCent_rms_MaxSpec 1.720e+03

The values for both the methods are high in general as the estimated f0 is one octave higher than the annotated for both the cases. As can be seen, the errCent_rms_ACF value is always greater than errCent_rms_ACF as there are lot of high spikes in between the contour as can been in Figure 5. These spikes increase the rms error but the estimated pitch is better in case of the autocorrelation based method. The one octave shift is probably because of the complex nature of the signal as opposed to the synthetic signals tested in problem 1 and 2. These errors might be due to the high magnitude at formant locations.

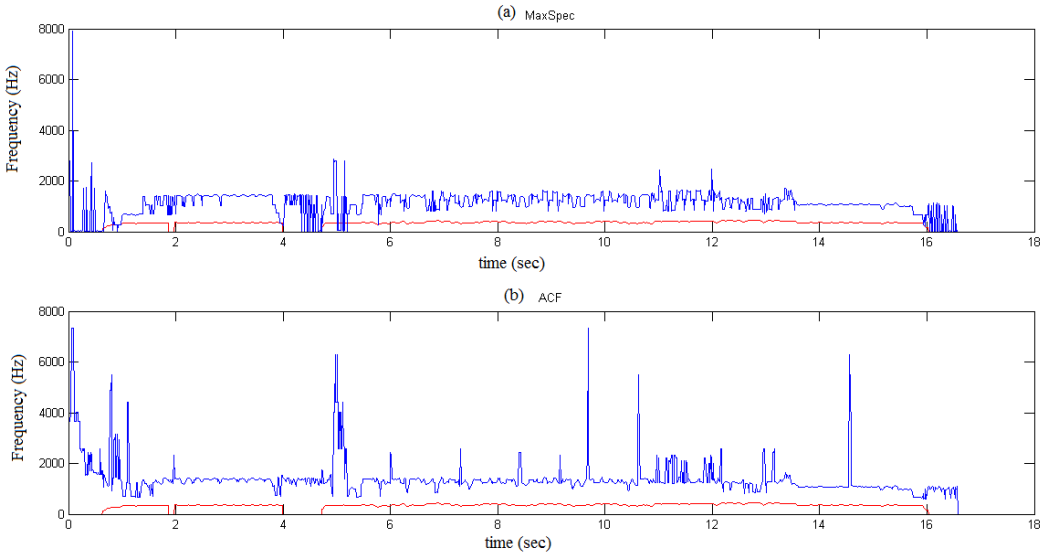

Figure 5. The red contour is the annotated f0 and the blue contour in (a) is estimated f0 using maximum spectral peak method while in (b) it is estimated using autocorrelation based method. The contours are for the test clip named '63-M2_AMairena.wav'

Potential solutions to improve are using a moving average filter and keeping the frequency values which don't show a sudden jump or don't fluctuate quickly. The current pitch value could be compared with the previous values to achieve this.

4. [5] (Bonus) Improvement

Implement a matlab function: [f0, timeInSec] = myPitchTrack_Mod(x, windowSize, hopSize, fs) that modifies (or combines) the above methods and provides better estimations. Please explain your approach in the report. Your function will be tested using a testing set (not provided), and points will be given based on its performance (comparing to the other two methods).

Since the pitch contour is erratic, a smoothing filter (moving Average) was applied to get a fine contour. Also, when we compare both the estimates, we observe that when there is a silence or sudden change in the signal, the Maximum Spectrum estimate is more desirable than the ACF estimation. Figure 6 highlights such differences.

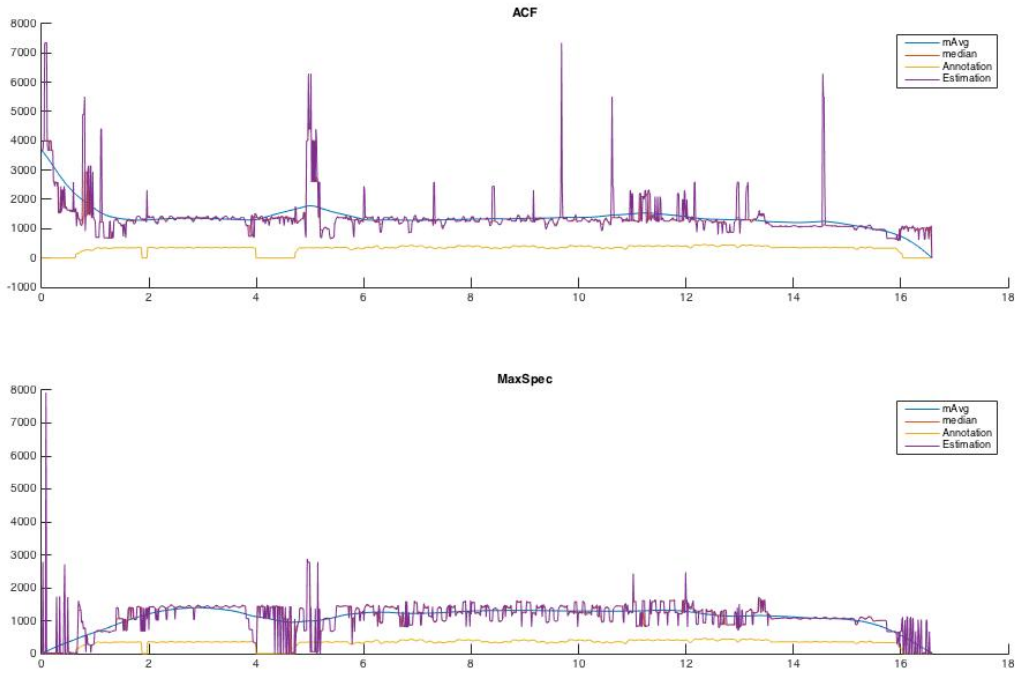

Figure 6. Moving median and mean filter modifications to the estimates. The contours are for the test clip named '63-M2_AMairena.wav'

We optimize by taking the minimum contour between the moving average estimates of Maximum Spectrum and Autocorrelation methods. The result is shown in Figure 7.

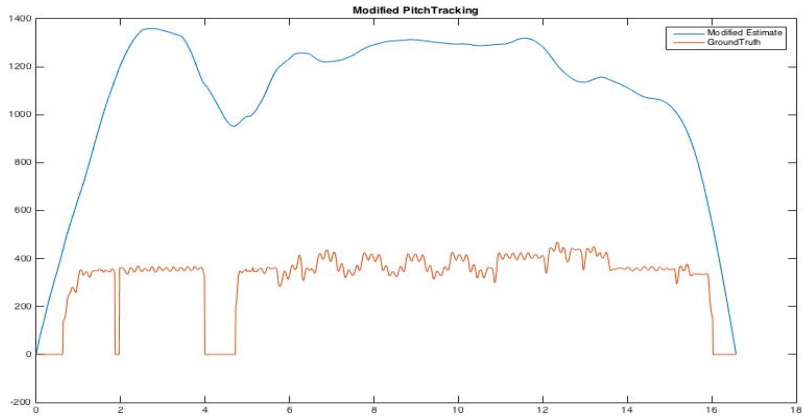

Figure 7. Pitch contour using the modification. The contours are for the test clip named '63-M2_AMairena.wav'

The error for the modified function is also less than both the methods for the audio in consideration as seen below:
63-M2_AMairena: errCent_rms =2.0130e+03

Another example in figure 8 of '01-D_AMairena.wav' file shows comparison of all the 3 extracted pitches and the modified pitch looks better than either of the methods individually.

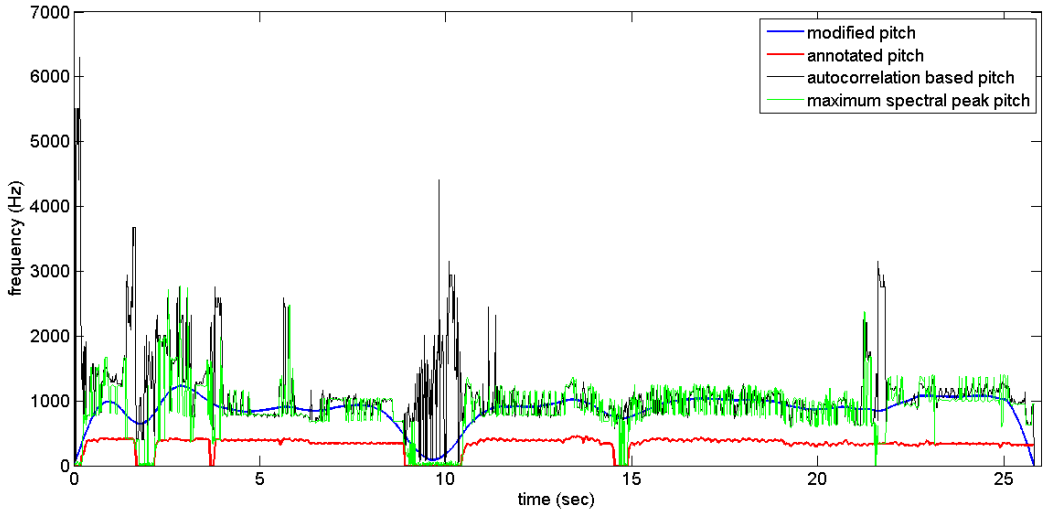

Figure 8. Plots of all the methods of f0 extraction with the ground truth. The modified methods seems to be better than either of the methods.