MATLAB Exercise – Autocorrelation Pitch Detector

Program Directory: `matlab_gui\autocorrelation_pitch_detector`
Program Name: `ac_pitch_GUI25.m`
GUI data file: `ac_pitch.mat`
Callbacks file: `Callbacks_ac_pitch_GUI25.m`
TADSP: Section 10.5.4, pp. 620-623, Problem 10.7

This MATLAB exercise implements a pitch period detection algorithm using the modified autocorrelation method with a specified center clipping threshold.

Autocorrelation Pitch Detector – Theory of Operation

This MATLAB exercise analyzes a designated speech utterance on a frame-by-frame basis, using a center-clipped, modified correlation analysis method. The algorithm finds the peak value of the modified correlation, within a specified pitch period interval, (different pitch period ranges are used for male and female talkers), computes a confidence score based on the autocorrelation level at the detected peak value compared to the autocorrelation level at zero lag, and decides whether a valid pitch period peak is detected in each frame. If the normalized peak value of the autocorrelation at the detected pitch period exceeds a user-specified threshold, that pitch period is designated as the pitch period at the given frame; otherwise the pitch period is designated to be 0 samples. A median smoother attempts to non-linearly interpolate the pitch period contour to account for frame errors within the detected pitch period contour.

Autocorrelation Pitch Detector – GUI Design

The GUI for this exercise consists of two panels, 4 graphics panels, 1 title box and 13 buttons. The functionality of the two panels is:

1. one panel for the graphics display,
2. one panel for parameters related to the signal processing for autocorrelation analysis, and for running the program.

The set of four graphics panels is used to display the following:

1. the waveform for the designated speech utterance,
2. the raw pitch period contour as estimated by the autocorrelation pitch detector,
3. the median-smoothed pitch period contour,
4. the confidence score for the pitch period estimates.

The title box displays the information about the selected file along with the set of autocorrelation analysis parameters. The functionality of the 13 buttons is:

1. a pushbutton to select the directory with the speech file that is to be analyzed using short-time analysis methods; the default directory is 'speech_files',
2. a popupmenu button that allows the user to select the speech file for analysis,
3. a pushbutton to play the designated speech file,
4. an editable button that displays the designated speech file sampling rate, f_s,
5. a popupmenu that specifies the gender of the talker for the designated speech utterance in order to restrict the range of pitch period estimates to those suitable for the gender of the talker,

6. an editable button that specifies the signal processing rate, `fsd`, of the autocorrelation pitch detector; (the default value is `fsd=6000` samples per second),
7. an editable button that specifies the frame duration, L_m, (in msec) for short-time analysis; (the default value is $L_m = 40$ msec),
8. an editable button that specifies the frame shift, R_m, (in msec) for short-time analysis; (the default value is $R_m = 10$ msec),
9. an editable button that specifies the center clipping level, `alpha`, as a percentage of the peak speech values in each analysis frame; (the default value of `alpha` is 60%),
10. an editable button that specifies the length of the median smoother, Lmed (in frames); (the default median smoother length is 5 frames),
11. an editable button that specifies the confidence threshold on autocorrelation values to accept the peak as the pitch period for the current frame; (the default value of the confidence threshold is 0.33),
12. a pushbutton to run the code and display the results of short-time analysis on the four graphics panel displays,
13. a pushbutton to close the GUI.

Autocorrelation Pitch Detector – Scripted Run

A scripted run of the program 'ac_pitch_GUI25.m' is as follows:

1. run the program 'ac_pitch_GUI25.m' from the directory 'matlab_gui\ac_pitch',
2. hit the pushbutton 'Directory'; this will initiate a system call to locate and display the filesystem for the directory 'speech_files',
3. using the popupmenu button, select the speech file for short-time feature analysis; choose the file 'test_16k.wav' for this example; the exercise then enters the original speech signal sampling rate in the designated button,
4. hit the pushbutton 'Play Speech File' to play the original speech file,
5. using the popupmenu button, select the gender of the talker as male or female; the gender for the designated speech file is male,
6. using the editable buttons, choose initial values of 10000 samples per second for the signal processing sampling rate, `fsd`; choose the value of 40 msec for the short-time analysis frame length, L_m; choose a value of 10 msec for the short-time analysis frame shift, R_m; choose a value of 60 for the center clipping percentage level, `alpha`; choose a value of 5 (frames) for the median smoother length, `Lmed`; and choose a value of 0.33 for the confidence threshold, `conf_thr`;
7. hit the 'Run AC Pitch Detector' button to compute and display the signal waveform in the top graphics panel; the raw pitch period contour in the second graphics panel; the median-smoother pitch period contour in the third graphics panel; and the confidence score in the bottom graphics panel,
8. experiment with different choices of speech file, and with different values for the parameters `fsd`, L_m, R_m, `alpha, Lmed` and `conf_thr`,
9. hit the 'Close GUI' button to terminate the run.

An example of the graphical output obtained from this exercise using the speech file 'test_16k.wav' is shown in Figure 1. The graphics panels show the speech waveform of the chosen starting frame (top graphics panel), the raw pitch period contour (second graphics panel), the median-smoothed pitch period contour (third graphics panel), and the confidence score (bottom graphics panel). It can be seen that most (but not all) of the single and double frame pitch period errors are fixed by the median smoother.

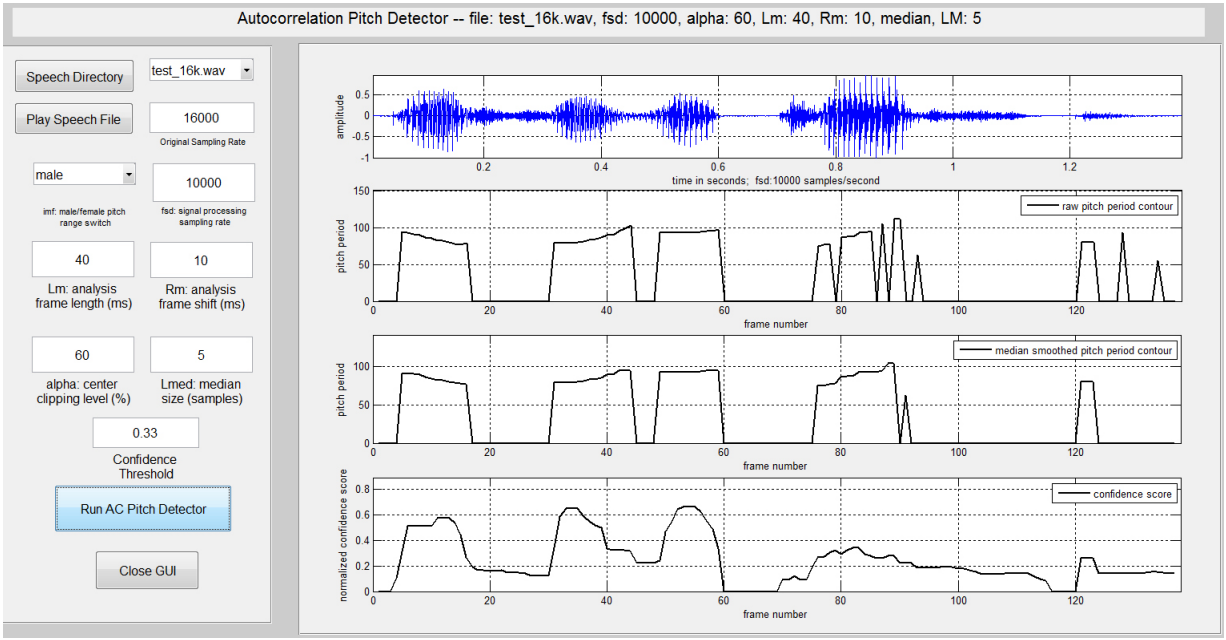

Figure 1: Graphical output from autocorrelation pitch period analysis of utterance 'This is a test', spoken by a male talker, The upper graphics panel shows the waveform, the second graphics panel shows the raw pitch period contour, the third graphics panel shows the median-smoothed pitch period contour, and the bottom graphics panel shows the confidence score for pitch period estimates.

Autocorrelation Pitch Detector – Issues for Experimentation

1. run the scripted exercise above, using the speech file 'test_16k.wav', and answer the following:

   - what is the range of the pitch period estimates (in both samples and in msec)?
   - what is the range of the pitch frequency estimates (in Hz)?
   - is the range of pitch frequency estimates consistent with the talker being male?
   - how often are there gross jumps in the raw pitch period contour? what explains these jumps?
   - how effective is the median smoother for this particular speech file set of pitch period estimates?
   - what is the perceptual impact of pitch period estimate jumps that are not smoothed out using the median smoother?
   - the pitch period estimates appear to be delayed with respect to the waveform plot but in reality they are perfectly aligned. Can you account for why they appear to be delayed (think of the filtering and start-up issues for voiced speech signals),
   - Compare the smoothed pitch period contour in the third plot to the raw pitch period data in the second plot. Does the median smoother do the correct thing all the time? Adjust the median length to remove pitch period errors that remain when the median length is too short; e.g., Lmed=3.

2. run the scripted exercise above, using the speech files 'we were away a year ago_lrr.wav', 'we were away a year ago_suzanne.wav', (remember to set the pitch range switch to 'female'), and 'vowels_100Hz_edited.wav',

3. how well does this pitch detector work on the above set of three different sentences?

4. are there any significant differences between male and female pitch period estimates (e.g., is one easier to estimate than the other); what does the synthetic vowel pitch period contour tell you about the theoretical performance on synthetic speech?

5. using the utterance 'test_16k.wav' vary the signal processing parameters, namely `fsd`, the signal processing sampling rate, L_m, the signal frame size in msec, R_m, the signal frame shift in msec, `alpha`, the center clipping percentage level, and finally `Lmed`, the duration of the median smoother in frames. Do the results show big changes with small changes of any of these signal processing parameters? (Recall that the signal processing parameter value can be moved up or down so don't just restrict your experimentation to upward or downward changes.)