MATLAB Exercise – Autocorrelation Estimates

Program Directory: `matlab_gui\autocorrelation_estimates`
Program Name: `autocorrelation_estimates_GUI25.m`
GUI data file: `autocorrelation_estimates.mat`
Callbacks file: `Callbacks_autocorrelation_estimates_GUI25.m`
TADSP: Section 6.5-6.6, pp. 265-275, Problem 6.20

This MATLAB exercise computes four types of short-time autocorrelation analysis of a selected speech frame and, in cases that are determined to represent voiced speech frames, estimates the pitch period of the current frame for each of the four types of short-time autocorrelation analysis. The exercise also processes the selected speech signal on a frame-by-frame basis, and estimates and plots the pitch period contour of the entire utterance.

Autocorrelation Estimates – Theory of Operation

From a user-designated speech file, this MATLAB exercise computes the short-time autocorrelation of the speech, on a frame-by-frame basis, using four types of short-time autocorrelation analysis discussed in TADSP, namely:

1. standard short-time autocorrelation analysis
2. modified short-time autocorrelation analysis
3. modified short-time, center-clipped, autocorrelation analysis
4. modified short-time, 3-level, autocorrelation analysis

In order to perform short-time autocorrelation analysis, the analysis frame length, L_m (specified in msec and converted to samples at the sampling rate of the speech file), the analysis frame shift, R_m (specified in msec and converted to samples at the sampling rate of the speech file), the analysis window (Hamming or rectangular), and the maximum number of autocorrelation lags, `pdhigh`, the minimum number of autocorrelation lags, `pdlow`, all need to be specified by the user. In order to compute each of the four types of short-time autocorrelation, the user also needs to specify an appropriate center clipping level (`P-Clipping %`) that will preserve the key features needed for reliable pitch period detection.

The exercise computes and plots the specified set of four autocorrelations, for each analysis frame of speech on a set of 8 graphics panels, showing both the analysis frame and the resulting short-time autocorrelation, for each of the four autocorrelation analysis methods. For each frame of voiced speech (i.e., frames that exceed some reasonable log energy threshold), the exercise finds the autocorrelation lag (within the allowable region of pitch periods), where the maximum autocorrelation occurs and marks this location as the pitch period estimate for the current analysis frame. By performing this pitch period estimation for each of the four methods of autocorrelation analysis, and for each frame of speech, the exercise estimates the pitch period contour for the entire speech signal.

Interactive Method of Defining the Speech Analysis Frame Starting Sample

Several MATLAB Exercises rely on frame-based analysis methods where the user needs to specify both the speech file for analysis, and the starting sample of the speech analysis frame of interest. The method that we have chosen to define the frame starting sample is an interactive analysis which homes in on an appropriate analysis frame in a series of steps. The operations of this interactive method for determining the starting sample of the speech analysis frame for autocorrelation analysis proceed as follows:

1. In a specified graphics frame (or figure sub-frame) a single line plot of the entire speech waveform is obtained, as illustrated at the top panel of Figure 1. A graphics curser then appears allowing the user to move the cursor to the region of speech that is of interest for specifying the current analysis frame. A solid vertical cursor is shown at the place selected by the user. For the example of Figure 1 the cursor location is approximately sample 13000, as indicated by the solid red bar.

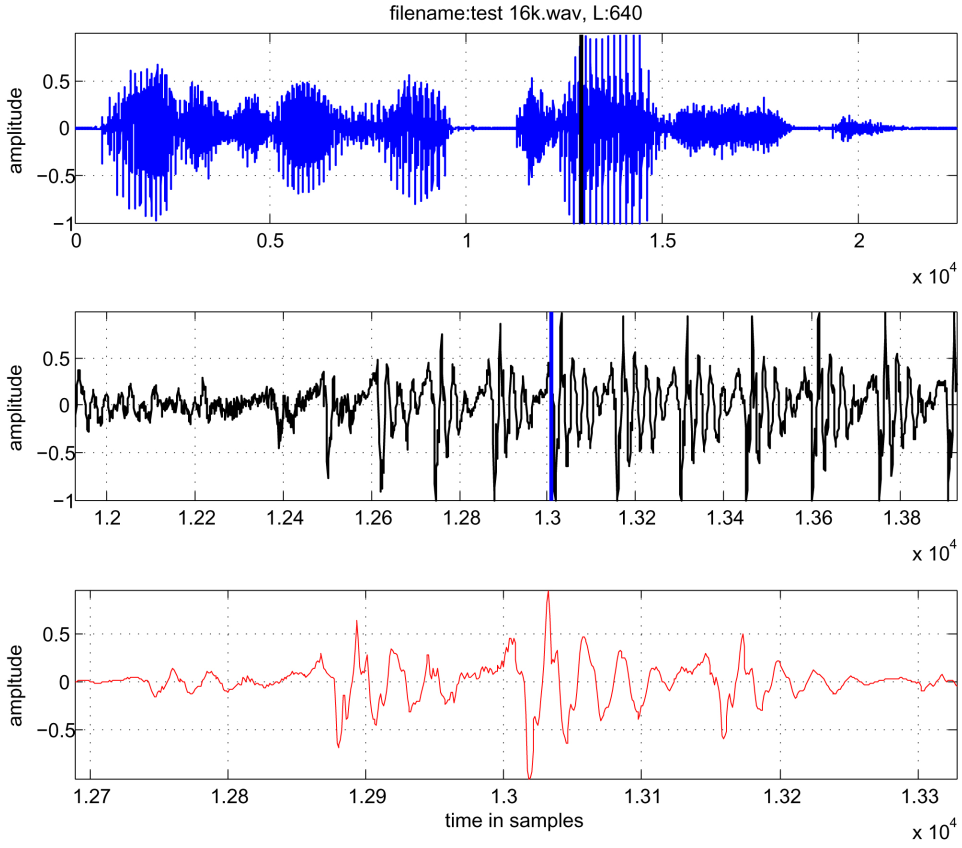


Figure 1: Sequence of waveform plots defining how the user can interactively choose a starting sample for the current analysis frame.

2. In another specified graphics frame (or figure sub-frame) a plot of the speech signal over a region that is about ± 1000 samples around the location of the cursor in the previous step; i.e., from sample 12000 to sample 14000. A second graphics cursor appears allowing the user to move the cursor to the exact starting sample of interest (to within the resolution of the display) for specifying the current analysis frame, as illustrated in the middle graphics panel of Figure 1. Here the cursor is again shown in the area of sample 13000.

3. The current analysis frame is then defined as the frame of speech from the starting sample of step 2 minus half the window length, to the starting sample of step 2 plus half the window length. The designated analysis frame is then weighted by the analysis window (Hamming in the case here) and plotted in the bottom graphics panel.

It should be clear that the three steps of the above process for choosing an analysis frame can be implemented in either a single graphics panel or frame (by simply overwriting the graphics panel with the new speech signal) or in a series of graphics panels or frames. The current exercise uses one of the 8 graphics panels and overwrites the speech waveform plot at each step of the analysis. This process is a very useful and efficient one for choosing a region of interest within the speech signal, and then homing into a particular analysis frame using the steps outlined above.

Autocorrelation Analysis – GUI Design

The GUI for this exercise consists of two panels, 8 graphics panels, 1 title box and 15 buttons. The functionality of the two panels is:

1. one panel for the graphics display,
2. one panel for parameters related to the computation of the four types of autocorrelation analysis, and for running the program.

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

1. the window-weighted speech section used for autocorrelation analysis (top left graphics panel),
2. the resulting short-time autocorrelation of the speech section (top right graphics panel),
3. the speech section used for modified autocorrelation analysis (a rectangular window is assumed here) (second left graphics panel),
4. the resulting short-time modified autocorrelation of the speech section (second right graphics panel),
5. the center clipped section of speech used for modified autocorrelation analysis (third left graphics panel),
6. the resulting short-time modified, center clipped, autocorrelation of the speech section (third right graphics panel),
7. the 3-level, center clipped, section of speech used for modified autocorrelation analysis (bottom left graphics panel),
8. the resulting short-time modified, 3-level center clipped, autocorrelation of the speech section (bottom right graphics panel).

The title box displays the information about the selected file for short-time autocorrelation analysis, including the user-selected speech filename, the starting sample for autocorrelation analysis, the frame length, L, (in samples) and frame shift, R, (in samples), the maximum lag (maxlag) for pitch detection, and the clipping level (P-Clipping %). The functionality of the 15 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 allow the user to play the chosen speech file,
4. a popupmenu button that lets the user choose either a Hamming or rectangular window as the short-time analysis window for standard autocorrelation analysis; (default is Hamming window),
5. a text button that specifies the starting sample of the speech analysis frame for short-time autocorrelation analysis; the starting sample is specified through the use of the procedure described in Section ,
6. an editable button that specifies the duration of the frame, L_m, (in msec) for short-time analysis; (the default value is $L_m = 40$ msec),
7. an editable button that specifies the frame shift, R_m, (in msec), for short-time analysis; (the default value is $R_m = 10$ msec),
8. an editable button that specifies the shortest allowable pitch period, `pdlow`, (in msec); (the default is 3.5 msec),
9. an editable button that specifies the longest allowable pitch period, `pdhigh`, (in msec); (the default value is 12.5 msec),
10. an editable button that specifies the percentage clipping level for the center-clipped speech autocorrelation method; (the default is 60%),

11. a pushbutton to determine the single frame starting sample, `ss`, using the iterative method described in Section ; this starting sample defines the current frame of the speech signal,

12. a pushbutton to run the analysis code and display the signal processing results using the current frame of the speech signal; this button can be pressed and used as often as desired, changing one or more analysis parameters while keeping the frame starting sample the same,

13. a pushbutton to run the analysis code and display the signal processing results using the next frame of signal; i.e., the frame with starting sample set to `ss+R` where `R` is the frame shift in samples; this button can be pushed repeatedly to provide a frame-by-frame analysis,

14. a pushbutton to run the code iteratively from the beginning of the file (the first frame) to the end of the file (the last frame) and create a contour of pitch period values over the duration of the signal, and display the resulting pitch period contour along with a confidence measure on a separate graphics figure,

15. a pushbutton to close the GUI.

Autocorrelation Analysis – Scripted Run

A scripted run of the program 'autocorrelation_estimates_rev1_GUI25.m' is as follows:

1. run the program 'time_domain_features_GUI25.m' from the directory 'matlab_gui\autocorrelation_estimates',

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 'we were away a year ago_suzanne.wav' for this example,

4. using the popupmenu button, choose Hamming for the short-time analysis window,

5. using the pushbutton 'Play Speech File' to play the chosen speech file,

6. the button which is labeled 'Starting Sample' is a text button which displays the starting sample of the user selected speech analysis frame; the process used to select the starting sample of the analysis frame was explained above,

7. using the editable buttons, choose the initial value of 40 msec for the frame length, L_m, and 10 msec for the frame shift, R_m,

8. using the editable buttons, choose the initial value of 3.5 msec for `pdlow` the shortest pitch period, and 12.5 msec for the longest pitch period, `pdhigh`,

9. using the editable button, choose a value of 60% for the center-clipping level percentage,

10. hit the 'Get Frame Starting Sample' button to interactively choose the initial analysis frame starting sample, `ss`, using the iterative method described above; try to choose the starting sample as close to the value of 1284 so as to match the plotted results for this example exercise,

11. hit the 'Run Current Frame' button to initiate single frame analysis of the speech beginning at the current frame starting sample, `ss`; the results of autocorrelation analysis using the four types of autocorrelation are shown in the various graphical plots; the 'Run Current Frame' button can be hit repeatedly after making changes in the analysis frame parameters,

12. hit the 'Run Next Frame' button to initiate single frame analysis on the next frame of speech, i.e., where the starting sample of the next frame is set to `ss+R`, where `R` is the frame shift in samples,

13. the single frame pitch period analysis algorithm estimates the pitch period for each of the four autocorrelation methods and indicates the value of the pitch period by a solid red line overlayed on each autocorrelation estimate; if no valid pitch period estimate is obtained, the pitch period defaults to the value 0,
14. the user can change any or all of the analysis parameters and redo the autocorrelation analysis from the same starting frame as used in the previous analysis; similarly the user can hit the 'Run Next Frame' button to increase the frame starting sample by R samples (where R is the frame shift in samples) and repeat the autocorrelation analysis on the new frame; the user can hit the 'Run Next Frame' button repeatedly, where each button push increases the starting sample of the frame analysis by R samples – i.e., a frame-by-frame or analysis can be performed in this manner,
15. hit the 'Run Pitch Detectors' button to sequentially perform autocorrelation analysis, beginning at sample 1, on each overlapping frame of speech, for each of the four autocorrelation methods, and thereby make a set of four estimates of the pitch period contour for the utterance; a confidence score is estimated for each frame and for each pitch detector, and this confidence score is used to smooth the pitch period contour and to eliminate estimation errors in periods of low certainty. The results of pitch detection for the entire utterance are shown in a separate figure (i.e., one not using the graphics panels),
16. experiment with different choice of speech file, and with different values for L_m, R_m, window type, and clipping level,
17. hit the 'Close GUI' button to terminate the run.

An example of the graphical output obtained from this exercise using the speech file:

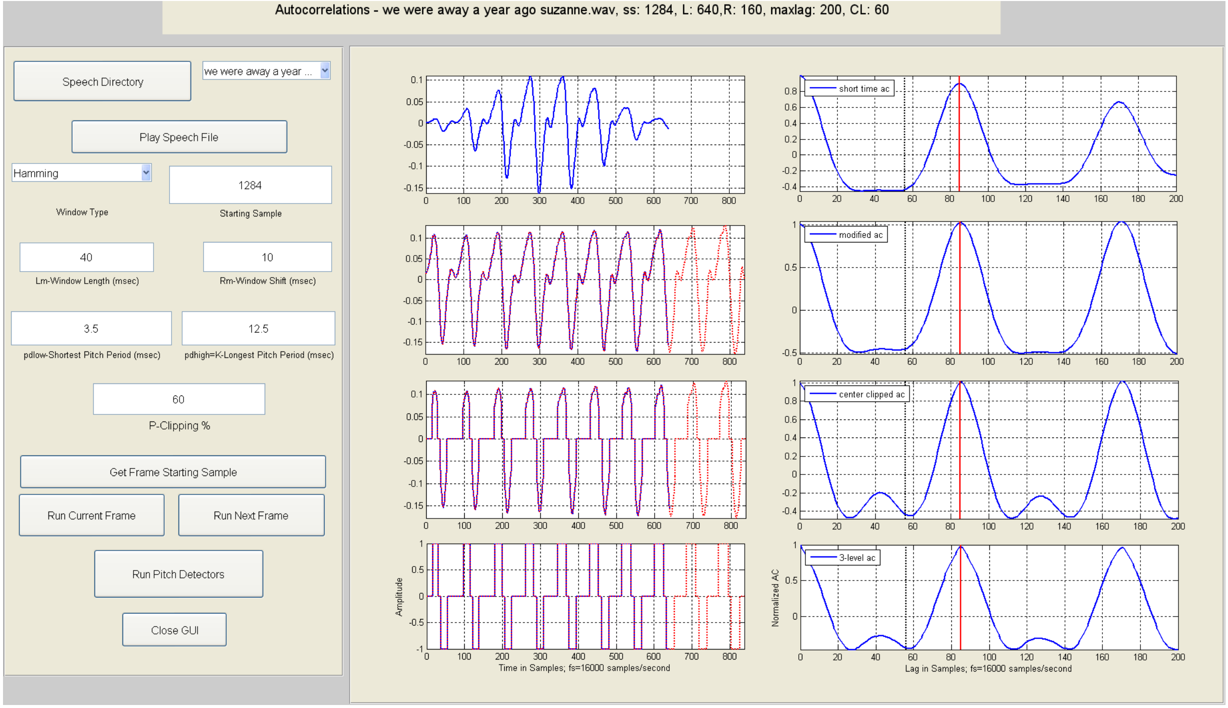
'we were away a year ago_suzanne.wav'

is shown in Figure 2. The graphics panels show the waveforms for the four autocorrelation analysis methods (on the left side of the display), along with the results of autocorrelation analysis (on the right side of the display). An estimate of the current frame pitch period, for each of the four pitch detectors, is indicated by the solid red vertical line in each autocorrelation plot.

A plot of the resulting pitch period contours for the entire utterance is shown in Figure 3 which shows the speech waveform (upper graphics panel), the set of four pitch period contours for each of the four methods of autocorrelation analysis (second graphics panel), the set of four confidence scores for each of the four methods of autocorrelation analysis (third graphics panel), and the short-time log energy contour of the utterance (fourth graphics panel).

Autocorrelation Analysis – Issues for Experimentation

1. run the scripted exercise above, and answer the following:
   - using the graphics cursor to select the middle sample of the frame to be as close to sample 1591 as possible so as to match the plot included in this manual; (recall that the starting sample is defined here as the middle sample of the frame minus half the frame length)
   - how similar are the waveform plots for the four types of autocorrelation analysis?
   - how similar are the autocorrelation plots for the four types of autocorrelation analysis?
   - what is the consensus pitch period (in samples) for this frame of speech?
   - given a sampling rate of $f_s = 16,000$ samples per second, what is the consensus pitch frequency in Hz?
   - what is the consensus value of the second largest peak in the four autocorrelation functions? Why isn't this secondary peak chosen as the pitch period for this frame of speech?
   - hold the frame position fixed and vary the frame length in an organized manner (e.g., in steps of 5 msec). Observe how the correlation peaks vary in amplitude among the different autocorrelation methods.



Autocorrelations - we were away a year ago suzanne.wav, ss: 1284, L: 640,R: 160, maxlag: 200, CL: 60

we were away a year ...

Window Type: Hamming

Starting Sample: 1284

Lm-Window Length (msec): 40

Rm-Window Shift (msec): 10

pdlow-Shortest Pitch Period (msec): 3.5

pdhigh=K-Longest Pitch Period (msec): 12.5

P-Clipping %: 60

Figure 2: Waveforms for the four autocorrelation analysis methods along with the resulting autocorrelation functions. The vertical red lines indicate the estimates of pitch period for each of the four methods of autocorrelation analysis.

- repeat the frame analysis with frame length fixed at 40 msec and with an organized set of clipping levels (e.g., from 10% to 90% in steps of 10%).
- observe what happens when the window length becomes about the same size as the pitch period.
- select the speech signal file '6B.waV' and run the pitch detectors with the default parameters. Which autocorrelation analysis methods appear to give the best results? By adjusting the window length, search boundaries, and clipping level, determine if it is possible to make all the pitch detectors give the same result, more or less.

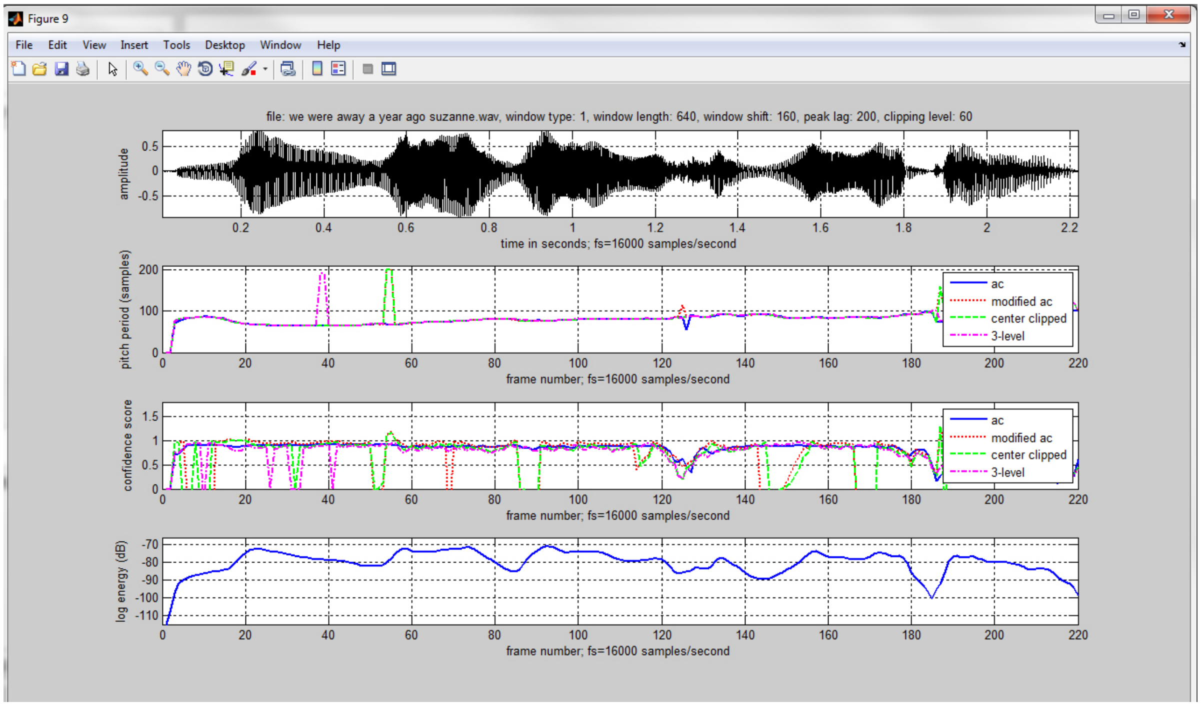
2. rerun the exercise choosing a completely different region of the speech waveform (consider using a section of speech where the middle sample of the frame is at sample 30644)

    - how different are the four autocorrelations from the previous starting sample?
    - what is the consensus pitch period for this frame of speech?
    - what are the main differences between the standard autocorrelation/modified autocorrelation and the center clipped autocorrelation/3-level autocorrelation?

3. change the value of the clipping level from 60 to 90 and run the exercise; then change the clipping level from 60 to 30 and run the exercise

    - what is the main change in the center-clipped autocorrelation with the change in clipping level
    - which clipping level (90, 60, or 30) best preserves the pitch peak with least distortion in the resulting autocorrelation estimate?

4. hit the 'Run Pitch Detector' button in order to perform autocorrelation analysis pitch detection using each of the four types of autocorrelation; using the new figure generated in this process, answer the following:



file: we were away a year ago suzanne.wav, window type: 1, window length: 640, window shift: 160, peak lag: 200, clipping level: 60

Figure 3: Speech waveform (upper plot) and four pitch period contours (second plot), and their confidence scores (third plot), and the speech log energy contour of the utterance (bottom plot).

- what types of errors in estimation of pitch period are made by the four autocorrelation methods
- which type(s) of autocorrelation appear to provide the smoothest and most error-free analysis?
- examining the confidence score panel, what can you say about the confidence of each of the four methods of autocorrelation analysis?