MATLAB Exercise – LPC Analysis With Varying p

Program Directory: `matlab_gui\lpc_varying_p`
Program Name: `lpc_varying_p_GUI25.m`
GUI data file: `lpc_varying_p.mat`
Callbacks file: `Callbacks_lpc_varying_p_GUI25.m`
TADSP: Section 9.4.3, pp. 497-500, Problem 9.27

This MATLAB exercise illustrates the matching behavior of the LPC polynomial log magnitude spectrum to the log magnitude spectrum of the STFT of a designated input frame of speech as a function of the LPC order, p, for a range of values of p; e.g., $p = 10, 20, 30, 40, 50$.

LPC Varying p – Theory of Operation

This MATLAB exercise processes a frame of a speech file (from a user-defined starting sample[1] and with a designated frame duration), and plots the LPC polynomial log magnitude spectrum as a function of LPC order, p, for each of 5 different values of p, and compare the results with the log magnitude spectrum of the STFT of the designated frame of speech.

LPC Varying p – GUI Design

The GUI for this exercise consists of two panels, 2 graphics panels, 1 title box and 14 buttons. The functionality of the two panels is:

1. one panel for the graphics display,
2. one panel for parameters related to the linear prediction analysis, and for running the program.

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

1. the speech frame, suitably weighted by the chosen window,
2. the log magnitude response of the speech frame STFT along with the LPC log magnitude spectrum from one of the three LPC analysis methods with 5 different values of LPC order, p.

The title box displays the information about the selected file along with the set of LPC analysis parameters. The functionality of the 14 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 current speech file,
4. a popupmenu button that specifies the LPC analysis method, `ilpc`; (the default is Autocorrelation, and the choices include Covariance and Lattice methods),
5. a popupmenu button that specifies the frame window for STFT analysis, `iwin_STFT`; (the default is a Hamming window),
6. a popupmenu button that specifies the frame window for LPC analysis, `iwin_LPC`; (the default is a Hamming window),
7. an editable button that specifies the initial LPC system order, `pinit`; (the default value is `pinit=10`),

[1] The iterative/interactive process for choosing the starting sample of the speech analysis frame is described in a later section.

8. an editable button that specifies the change in LPC system order, `pdelta`, for each of 5 iterations of LPC analysis with varying values for p; (the default value is `pdelta=10`),

9. an editable button that specifies the frame duration, L_m, (in msec) for short-time analysis; (the default value is $L_m = 40$ msec),

10. an editable button that specifies the frame shift, R_m, (in msec) for short-time analysis; (the default value is $R_m = 10$ msec),

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

12. a pushbutton to run the LPC 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 LPC 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 close the GUI.

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.

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.

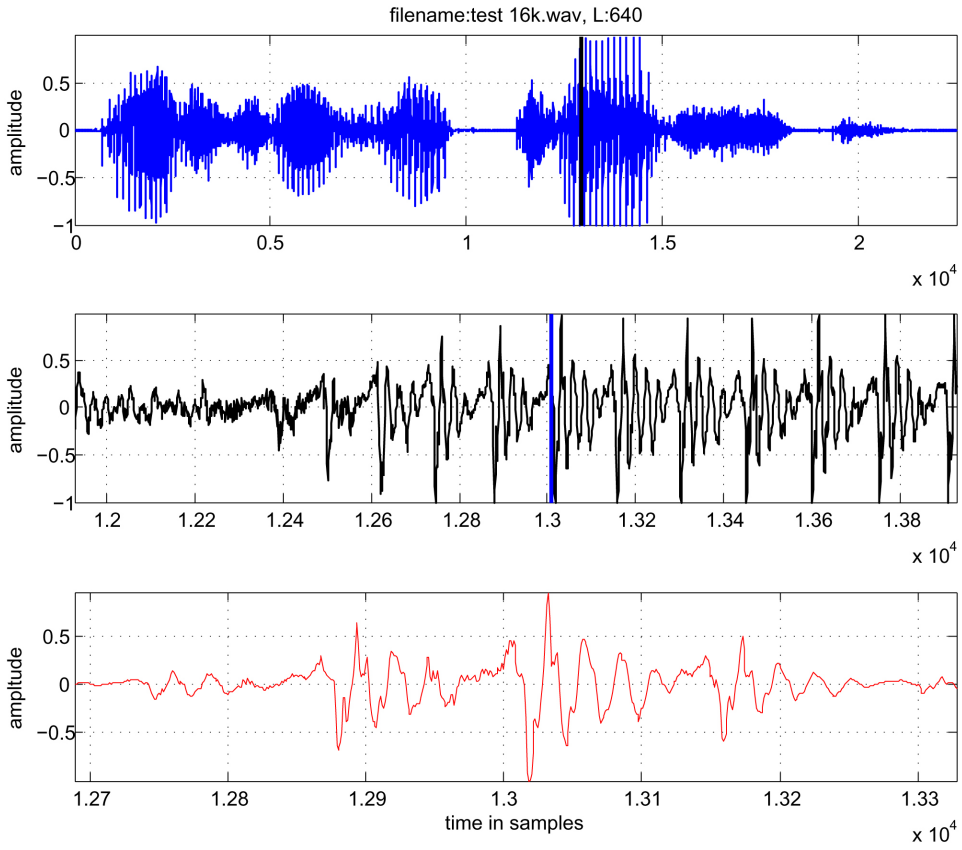

Figure 1: Sequence of waveform plots defining how the user can interactively choose a starting sample for the current analysis frame.

LPC Varying p – Scripted Run

A scripted run of the program 'lpc_varying_p_GUI25.m' is as follows:

1. run the program 'lpc_varying_p_GUI25.m' from the directory 'matlab_gui\lpc_varying_p_solution',

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. hit the pushbutton 'Play Speech File' to verify the contents of the selected speech file,

5. using the popupmenu buttons choose Autocorrelation method as the type of LPC analysis, `ilpc`; choose a Hamming window as the STFT window; and choose a Hamming window as the LPC Autocorrelation method analysis window,

6. using the editable buttons, choose a value of 10 for the LPC system order, `pinit`, and choose a value of 10 for the change in p per iteration, `pdelta`; choose a value of 40 msec for frame duration, L_m; and choose a value of 10 msec for frame shift, R_m,

7. hit the 'Get Frame Starting Sample' button to interactively choose the initial analysis frame starting sample, `ss`, using the iterative method described in the previous section; try to choose the starting sample as close to the value of 3132 so as to match the plotted results for this example exercise,

8. hit the 'Run Current Frame' button to initiate single frame LPC analysis of the speech beginning at the current frame starting sample, `ss`; the results of LPC analysis are shown in the graphical plot; the 'Run Current Frame' button can be hit repeatedly after making changes in the analysis frame parameters,

9. 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,

10. experiment with different choices of speech file, and with different values for L_m, R_m, `pinit, pdelta`, window types and LPC analysis method,

11. 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 speech frame used for STFT and LPC spectrum analysis (upper graphics panel), and the log magnitude spectral response of the STFT analysis along with the LPC log magnitude spectrum for each of the five values (`pinit, pinit+pdelta,...,pinit+4*pdelta`) of LPC system order (lower graphics panel). It can clearly be seen in Figure 2 that the accuracy of the LPC spectrum fit improves with increased values of p.

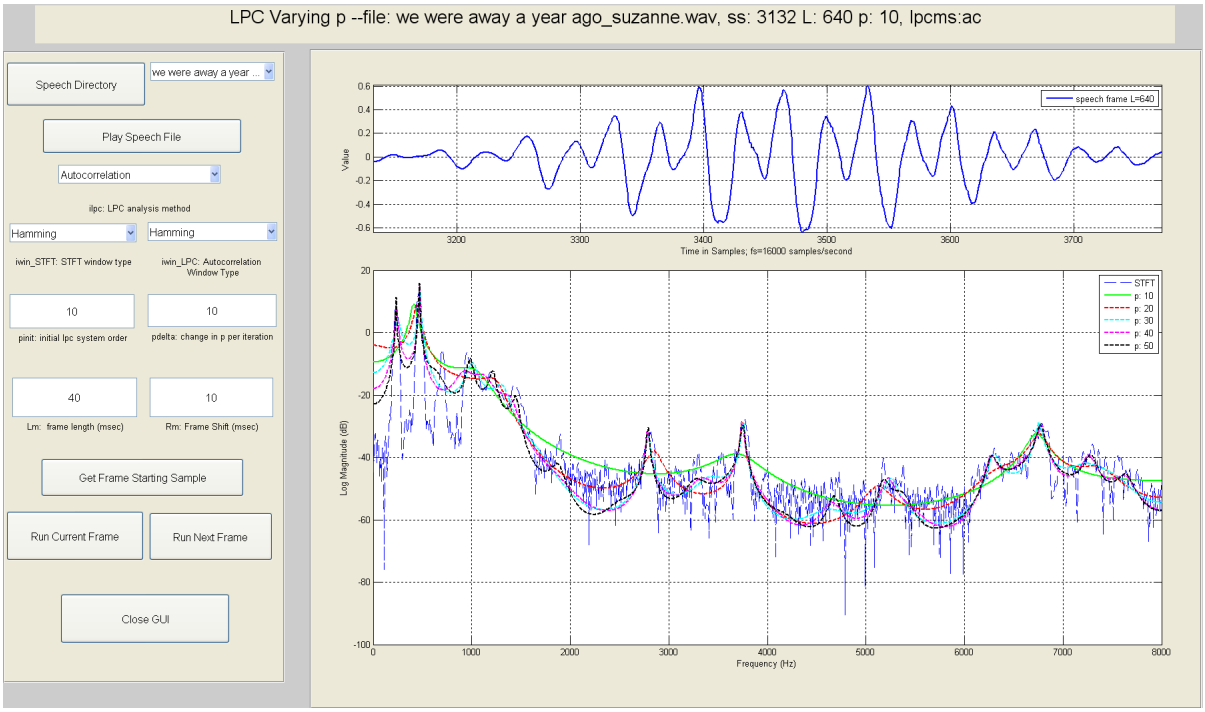

Figure 2: Plot of the speech frame (weighted by the selected window), in the upper graphics panel, and the log magnitude responses of the STFT of the speech frame along with the log magnitude responses of the LPC system with 5 different system orders, namely p =10, 20, 30, 40 and 50

LPC Analysis for Varying p – Issues for Experimentation

1. run the scripted exercise, above, using the speech file 'test_16k.wav' with frame starting sample, ss=1412 (or a starting sample that is reasonably close to ss=1412) and answer the following:

   - how closely do the LPC model spectra for the range of p values shown in the bottom graphical panel match the STFT spectrum or spectrum envelope (shown as the dashed curve in the bottom graphical panel)?
   - what accounts for the different degree of spectral envelope match as p varies from 10 to 50?
   - what are the two key ranges of p that determine the type of spectral match, namely what is the low range value of p (the one that matches the significant resonances in the STFT spectrum envelope), and what is the high range value of p (the one that matches the STFT spectrum namely the spectral peaks associated with the pitch harmonics)?

2. run the scripted exercise, above, this time using the synthetic vowel file 'vowels_100Hz_edited.wav', with frame starting sample ss=200.

   - what explains the behavior of the LPC spectral match to the STFT spectrum envelope as p varies from 10 to 50?
   - if the range of values of p is changed to the range $p = 2, 4, 6, 8, 10$ what changes in spectral match occur?
   - if the range of values of p is changed to the range $p = 100, 150, 200, 250, 300$ what changes in spectral match occur?
   - what explains the different quality of spectral match to the STFT spectrum and the STFT spectrum envelope for different ranges and values of p?

3. select a speech file and move to a region of voiced speech using the graphics cursor; run the STFT and LPC analyses using the default parameters. Compare the LPC log magnitude spectra for different values of p. Where do the peaks of the log magnitude response concentrate in the STFT log magnitude display?

4. where do the peaks of the LPC log magnitude spectra occur, relative to those of the STFT log magnitude spectrum, for the 5 different values of p?

5. what value of p gives the smoothest LPC log magnitude spectrum fit to the STFT log magnitude spectrum?

6. using the same frame of voiced speech as used in the three steps above, change to a rectangular window for both STFT and LPC analyses. What are the major effects of this window change (from Hamming to Rectangular) on the STFT log magnitude spectrum, and on the LPC log magnitude spectrums for the 5 values of p?

7. move to the next frame and observe the changes in the smoothed log magnitude spectrums from the previous frame. How different are the STFT and LPC log magnitude spectrums in adjacent speech frames? How much does this difference change as the frame shift parameter, R_m is increased from 10 msec to 20 or 30 msec?

8. repeat the steps of this exercise using an unvoiced region of speech. Are the spectral matching effects larger or smaller for unvoiced speech frames?