MATLAB Exercise – Spectral Smoothing Comparisons

Program Directory: `matlab_gui\spectral_smoothing`
Program Name: `spectral_smoothing_GUI25.m`
GUI data file: `spectral_smoothing.mat`
Callbacks file: `Callbacks_spectral_smoothing_GUI25.m`

This MATLAB exercise smooths the log magnitude spectrum of a frame of voiced speech using both homomorphic liftering and LPC analysis and compares and contrasts the resulting spectrums with that obtained from wideband spectrum analysis using the STFT.

Spectral Smoothing – Theory of Operation

This MATLAB exercise processes a designated frame of voiced speech from a user specified speech file, using a designated window and calculates the log magnitude spectrum of the STFT of the windowed frame of speech. Next the exercise does LPC analysis of the speech frame using a p^{th}-order predictor and computes the log magnitude spectrum of the LPC polynomial. Finally cepstral analysis is performed with the cepstrum liftered using a low quefrency lifter (with lowest retained quefrency specified by the user). Finally the exercise plots the three log magnitude spectrums on a common plot for comparison of the three log magnitude spectrums.

Spectral Smoothing – GUI Design

The GUI for this exercise consists of two panels, 2 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 cepstral and the linear prediction analysis, and for running the program.

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

1. the current frame of speech used for STFT, cepstral and LPC analyses,
2. the log magnitude response of the speech frame STFT along with the log magnitude spectrum from the LPC analysis method and from the cepstral analysis method.

The title box displays the information about the selected file along with the set of LPC and cepstrum 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 current speech file,
4. an editable button that specifies the frame duration, L_m, (in msec) for short-time analysis; (the default value is $L_m = 40$ msec),
5. an editable button that specifies the frame shift, R_m, (in msec) for short-time analysis; (the default value is $R_m = 10$ msec),
6. an editable button that specifies the LPC system order, p; (the default value is $p = 16$),
7. an editable button that specifies the maximum quefrency for the low quefrency lifter, `Lcm` (default is `Lcm=4` msec),

8. a popupmenu button that lets the user choose either a Hamming or Rectangular window as the STFT analysis window; (default is Hamming window),
9. a popupmenu button that lets the user choose either a Hamming or Rectangular window as the LPC/cepstrum analysis window; (default is Hamming window),
10. 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,
11. 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,
12. 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,
13. a pushbutton to close the GUI.

Interactive Method of Defining the Speech Analysis Frame Starting Sample

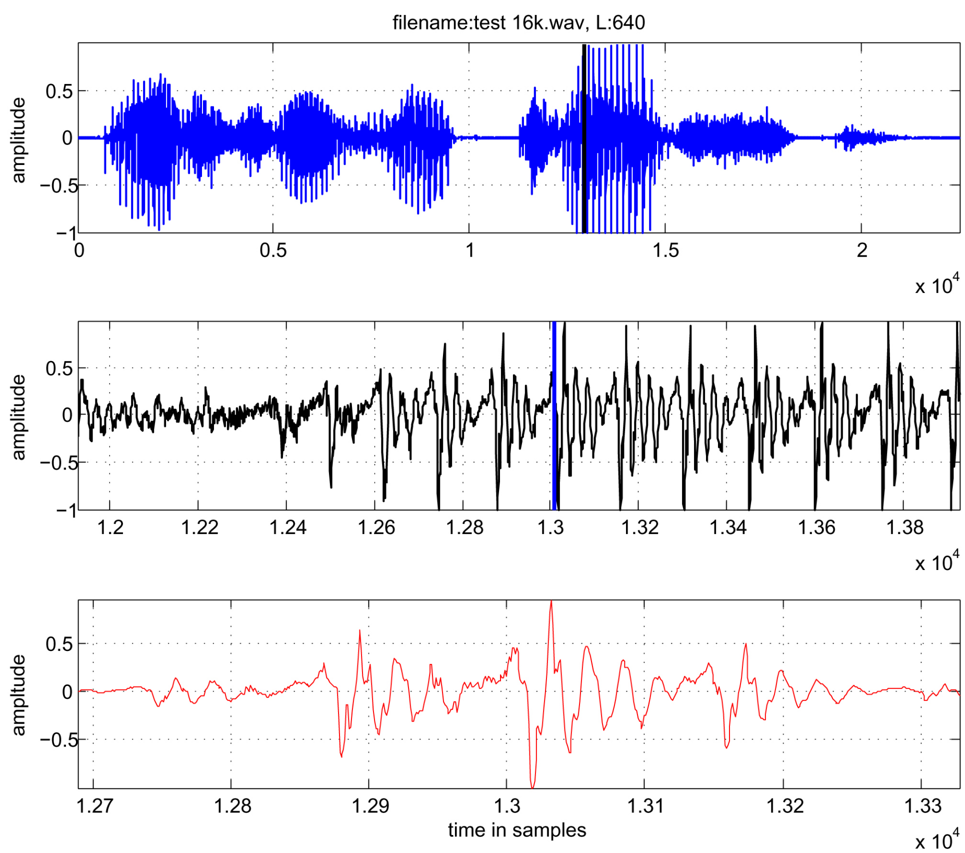

Figure 1: Sequence of waveform plots defining how the user can interactively choose a starting sample for the current analysis frame.

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.

Spectral Smoothing – Scripted Run

A scripted run of the program 'spectral_smoothing_GUI25.m' is as follows:

1. run the program 'spectral_smoothing_GUI25.m' from the directory 'matlab_gui\spectral_smoothing',

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_lrr.wav' for this example,

4. hit the pushbutton 'Play Speech File' to play the current file,

5. using the editable buttons, choose an initial value of 40 msec for the frame length, L_m; choose an initial value of 10 msec for the frame shift, R_m; choose an initial value of 12 for LPC system order, p; choose an initial value of 4 msec for the cepstral lifter cutoff quefrency, `Lcm`,

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

7. using the popupmenu button choose Hamming for the short-time LPC/cepstral window, `Lc`,

8. 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 7685 so as to match the plotted results for this example exercise,

9. 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 LPC and cepstral analysis are shown in the various graphical plots; the 'Run Current Frame' button can be hit repeatedly after making changes in the analysis frame parameters,

10. 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,

11. hit the 'Starting Sample for Current Frame' button to compute and display the current frame signal waveform in the upper graphics panel,

12. hit the 'Run Current Frame' button to perform spectral smoothing of the STFT frame of speech using cepstral and LPC analysis methods to get the smoothed spectral curves; the log magnitude responses of the STFT spectrum, the cepstrally smoothed spectrum, and the LPC estimated spectrum are all displayed in the lower graphics panel,

13. hit the 'Run Next Frame' button to update the current frame starting sample, and then run the spectral smoothing analysis on this updated frame of speech,

14. experiment with different choices of speech file, and with different values for L_m, R_m, p, `Lcm` and window types,

15. 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_lrr.wav' is shown in Figure 2. The graphics panels show the speech waveform (the current frame of speech) in the top graphics panel, and the log magnitude spectral response of the STFT analysis along with the LPC spectrum and the cepstrally smoothed spectrum (bottom graphics panel).

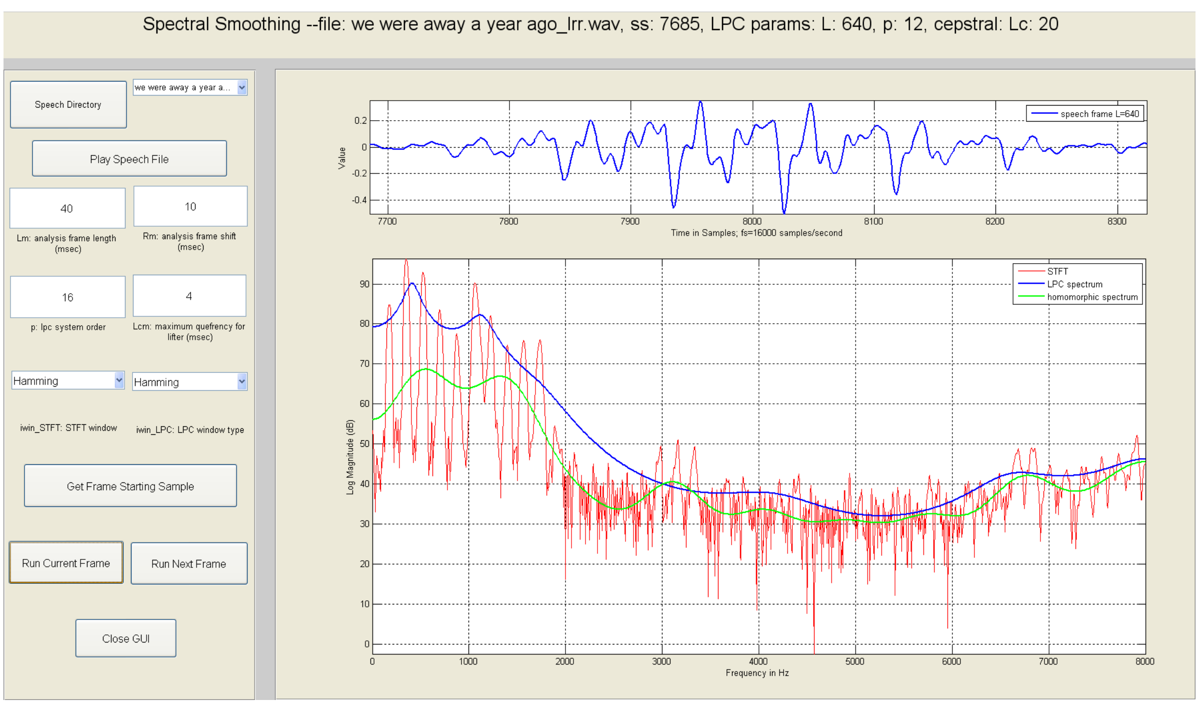

Figure 2: Graphical output from LPC analysis of a frame of voiced speech. The upper graphics display shows the speech waveform in the vicinity of the selected speech frame; the middle graphics panel shows the window-weighted speech signal; the bottom graphics panel shows the log magnitude response of the STFT of the speech frame along with the LPC spectrum and the cepstral analysis smoothed spectrum.

Spectral Smoothing – Issues for Experimentation

1. run the scripted exercise above, using the speech file 'we were away a year ago_lrr.wav' with starting sample ss=7473 (or some close starting sample), and answer the following:

   - what type of spectral analysis (narrowband or wideband) is performed in the STFT calculation?
   - what are the key differences between the LPC spectral fit and the low quefrency cepstral smoothing fit to the STFT spectral plot:
   - what happens to the LPC polynomial fit when the LPC system order, p, is increased to $p = 20$ or 24, or when the LPC system order, p, is decreased to $p = 8$?
   - what happens to the cepstral fit to the log magnitude spectrum when the low quefrency cepstral lifter uses either larger or smaller values of `Lcm` such as `Lcm=8` or `Lcm=2`?

2. repeat the scripted exercise above, using the synthetic speech file 'vowels_100Hz_edited.wav' with starting sample set to `ss=200` (or some close starting sample). How do the answers to the questions above change with this synthetic signal?

1. select a speech file and go to a frame of voiced speech as the current starting frame. Using the default parameters, run the scripted program. Compare the LPC and cepstral spectral estimates. Which spectral smoothing procedure seems to be doing the best job of extracting the formant resonances for this frame of speech?

2. Using the same frame of speech, increase and decrease the value of p, the LPC system order. What is the effect of these changes on the quality of the spectral matches from LPC analysis?

3. using the same frame of speech, increase and decrease the maximum quefrency of the lifter. What is the effect of these changes on the quality of the spectral matches from cepstral analysis?

4. set the LPC system order to 16, and the lifter quefrency to 25 and perform the spectral smoothing operation. Now move to the next frame and observe the changes in the smoothed spectra. How similar are the matches for this new frame?

5. move to a section of speech to an unvoiced region and repeat the above analyses. What is the major difference in the quality of the matches between voiced analyses and unvoiced analyses?