MATLAB Exercise – Formant Estimation from LPC Roots

Program Directory: `matlab_gui\formant_estimation`
Program Name: `formant_estimation_GUI25.m`
GUI data file: `formant_estimation.mat`
Callbacks file: `Callbacks_formant_estimation_GUI25.m`
TADSP: Section 10.6, pp. 635-645, Problem 10.11

This MATLAB exercise illustrates an algorithm for estimating the locations of the formants of voiced speech intervals based on the locations of the speech polynomial roots as obtained from a frame-based analysis of a speech signal using the method of Linear Prediction Analysis. A key aspect of this algorithm is finding regions of quasi-contiguous speech frames (where the first three formants form a contiguous region over the duration of the voiced region), and then extending these regions (both backward and forward) based on a somewhat weaker criterion for contiguous behavior of the formants.

Formant Estimation from LPC Roots – Theory of Operation

This exercise estimates the formants of voiced speech intervals based on the computed roots of a frame-based LPC analysis over the duration of a speech signal. The LPC roots are determined by factoring the LPC polynomial derived for each frame of signal, and by eliminating all real roots, all roots whose imaginary components are negative (these roots are represented by their complex conjugates whose imaginary components are positive), and all roots whose magnitude (radius in the z-plane) is below a user specified threshold. The resulting formant contours are estimated by finding contiguous regions with the first three formants of a given frame matching the first three formants of the previous frame within a specified matching criterion (based on the normalized distance between the first three formants of the pair of frames being matched). Finally a weaker criterion is used to extend the formant regions based on matches of only either the first two formants or based on matches of only the second and third formants.

Formant Estimation from LPC Roots – GUI Design

The GUI for this exercise consists of two panels, 3 graphics panels, 1 title box and 10 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 three graphics panels is used to display the following:

1. the speech waveform for the designated speech file,
2. the locations of the roots of the LPC polynomial for all frames of speech (only roots whose magnitude exceeds a threshold are plotted),
3. the contours of the first four formants based on the formant estimation algorithm.

The title box displays the information about the selected file along with the set of LPC analysis parameters. The functionality of the 10 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 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 minimum radius, `minrad`, for a root to be considered a candidate for a formant track; (the default value for `minrad` is 0.9),
8. an editable button that specifies a threshold on the normalized formant distance, `dthr`, between two tracks in order to align and merge the tracks; (the default value is `dthr=0.05`),
9. a pushbutton to run the code and display the results of LPC-based formant estimation on the two graphics panel displays,
10. a pushbutton to close the GUI.

Formant Estimation from LPC Roots – Scripted Run

A scripted run of the program 'formant_estimation_GUI25.m' is as follows:

1. run the program 'formant_estimation_GUI25.m' from the directory 'matlab_gui\formant_estimation',
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 designated speech file,
5. using the editable buttons, choose initial values of 40 msec for the analysis frame length, L_m; choose a value of 10 msec for the analysis frame shift, R_m; choose a value of 16 for LPC system order, p; choose a value of 0.9, for the minimum pole radius threshold, `minrad`; and choose a value of 0.05 for the formant distance threshold, `dthr`,
6. hit the 'Run Formant Estimation' button to compute and display the raw roots of the LPC polynomial, for each frame of speech, where the root radius exceeds the minimum radius threshold in the upper graphics panel, and to display the contours of the roots, as estimated by the algorithm, in the lower graphics panel,
7. experiment with different choices of speech file, and with different values for L_m, R_m, p, minrad and dthr,
8. 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 1. For this example the speaker is a male and the utterance consists of all voiced frames (with the exception of the /G/ in /ago/). The upper graphics panel shows the speech waveform, the middle graphics panel shows the set of LPC roots, on a frame-by-frame basis, where the root radius exceeded the threshold, minrad, and the lower graphics panel shows the contours of LPC roots (created from the roots of consecutive frames where the distance between roots in adjacent frames fell below the threshold, dthr).

A second example of LPC-based formant estimation is shown in Figure 2, again for the utterance 'we were away a year ago_lrr.wav'. For this example the speaker is a female and the utterance again consists of all voiced frames (with the exception of the /G/ in /ago/). Here it can be seen that the raw roots appear to preserve formant locations fairly well throughout the speech file; however from the formant contour plots, it is readily seen that the algorithm cannot create consistent and reliable formant estimates from the raw LPC roots, resulting in the plot of formant contours shown in the lower graphics panel.

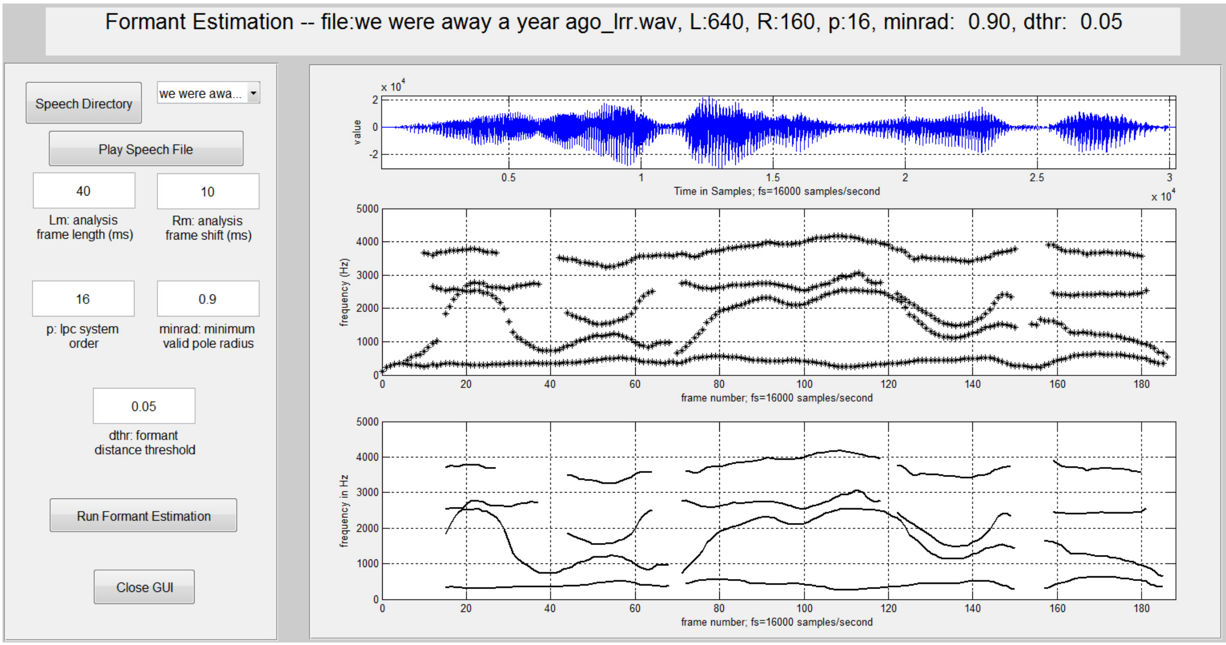

Figure 1: Graphical output from LPC formant estimation procedure. The upper graphics panel shows the speech waveform, the middle panel shows the set of LPC roots, for each frame of speech, where the root radius exceeded a minimum root radius threshold; the bottom graphics panel shows the contours of LPC roots (the formant contours) based on distances between roots of adjacent frames.

Formant Estimation from LPC Roots – Issues for Experimentation

- using one of the recording programs in this set of MATLAB speech processing exercises, record in separate output files the vowels IY, AA and UW and the diphthong OI. (Alternatively you can run this exercise using the speech files 'ih.wav', 'iy.wav', 'ah.wav', and 'oI.wav' that are available in the speech files directory). Use each of these files as input to the formant estimation procedure. Do the formants that are estimated for these vowels agree with the acoustic theory for the location of these vowels? Can you see the vowel transition in the formants of the diphthong 'oI.wav'?

- using the speech file 'we were away a year ago_lrr.wav', observe the effect of changing the minimum pole radius parameter to values that are smaller than the default (e.g., `minrad=0.8` and 0.85), or to values that are larger than the default (e.g., `minrad=0.95`). What impact does this have on the LPC roots, and the attempt to organize the roots into smooth formant tracks?

- repeat the previous analysis but this time consider changing only the LPC analysis order. What happens to the LPC roots when p is decreased to values such as 10, 12 or 14? What happens to the LPC roots when p is increased to values such as 18 or 20?

- using the speech file '9A.waV', observe the two nasal regions of the speech utterance /nine/. Are the formants clear for the nasal regions at the beginning and end of the utterance? Are the vowel regions also clear for this utterance?

- using the speech file '6A.waV', observe the fricative regions at the beginning and end of the speech utterance /six/. How should the minimum radius be set to automatically remove most of the roots for the fricative regions?

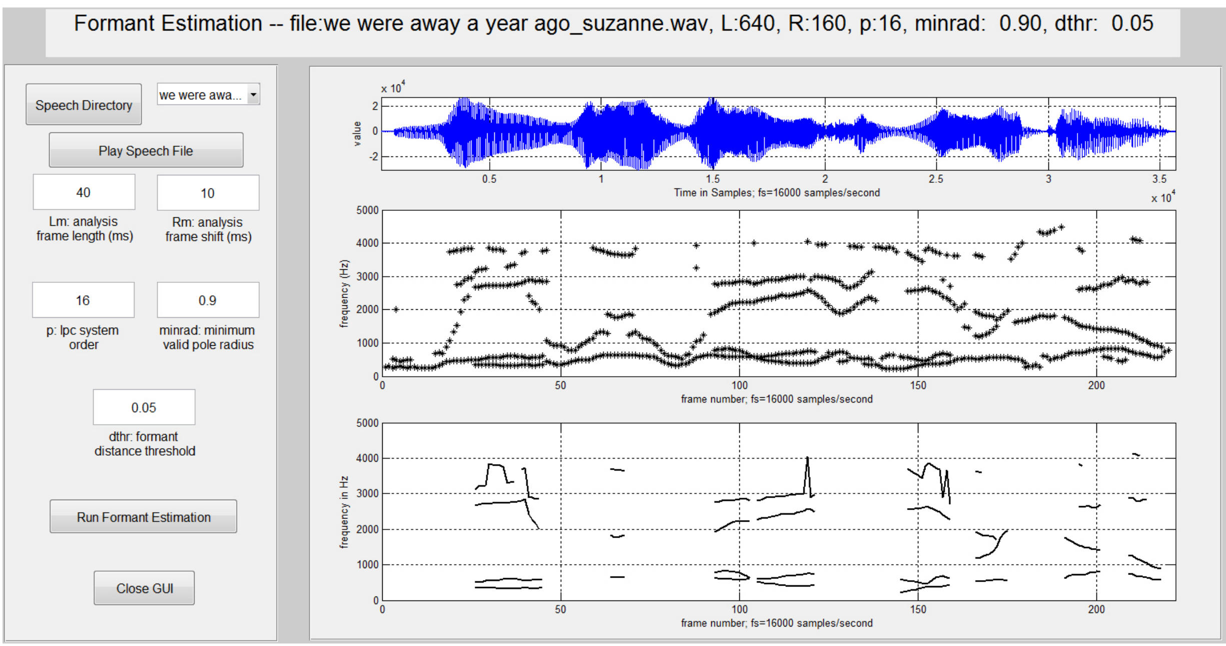

Figure 2: Graphical output from LPC formant estimation procedure. The upper graphics panel shows the speech waveform, the middle graphics panel shows the set of frame-based LPC roots, for each frame of speech, where the root radius exceeded a minimum root radius threshold; the bottom graphics panel shows the contours of LPC roots (the formant contours) based on distances between roots of adjacent frames.