MATLAB Exercise – Cepstral Liftering of Speech

Program Directory: `matlab_gui\cepstral_liftering`
Program Name: `cepstral_liftering_GUI25.m`
GUI data file: `cepstral_liftering.mat`
Callbacks file: `Callbacks_cepstral_liftering_GUI25.m`
TADSP: Section 8.5, pp. 440-454, Problem 8.18

This MATLAB exercise shows the spectral smoothing effects of low quefrency cepstral liftering on the log magnitude spectrum of a speech signal. This is done by first computing the real cepstrum of a window-weighted frame of speech and saving the resulting log magnitude spectrum as the baseline (unsmoothed) speech spectrum. A low quefrency lifter is then used to effectively smooth the log magnitude spectrum, with the cutoff quefrency varied from a low value of 20 to a high value of 100 in steps of 20 quefrencies. The most smoothing occurs with low cutoff quefrencies for the lifter; the least smoothing occurs with high cutoff quefrencies for the lifter.

Cepstral Liftering of Speech – Theory of Operation

Using a designated frame of speech from a given speech file, of specified length and specified window weighting function, the exercise computes the real cepstrum of the speech frame using the conventional DFT method. Next the exercise computes the log magnitude frequency response of the speech frame and uses it as a baseline for comparison. Next the exercise lifters the real cepstrum using a low quefrency lifter (with specified cutoff quefrency) and computes the resulting (cepstrally) smoothed log magnitude spectrum. This liftering operation is repeated using different lifter quefrency cutoffs (20, 40, 60, 80 and 100 quefrencies) enabling the user to examine the effects of various length low quefrency lifters on the (cepstrally smoothed) log magnitude spectrum of the speech signal.

Cepstral Liftering of Speech – GUI Design

The GUI for this exercise consists of two panels, 3 graphics panels, 1 title box and 12 buttons. The functionality of the two panels is:

1. one panel for the graphics display,
2. one panel for parameters related to short-time cepstral analysis, and for running the program.

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

1. the current speech frame used for cepstral analysis,
2. the real cepstrum of the current frame of speech, with the zeroth cepstral component set to 0,
3. log magnitude frequency response of the current speech frame and the cepstrally smoothed log magnitude frequency response using low quefrency lifters of length 20, 40, 60, 80 and 100 samples (quefrencies).

The title box displays the information about the selected file and related signal processing parameters for analysis of short-time cepstrums. The functionality of the 12 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 enable the user 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. a popupmenu button that lets the user choose either a Hamming or rectangular window as the short-time analysis window; (default is Hamming window),

7. an editable button that specifies the length (in msec) of the cepstrum, `ceplm`, to be computed and displayed; (default is `ceplm=30` msec),

8. an editable button that specifies the FFT size, `nfft`, for DFT-based cepstral analysis; (default is `nfft=2048`),

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

10. 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,

11. 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,

12. 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.

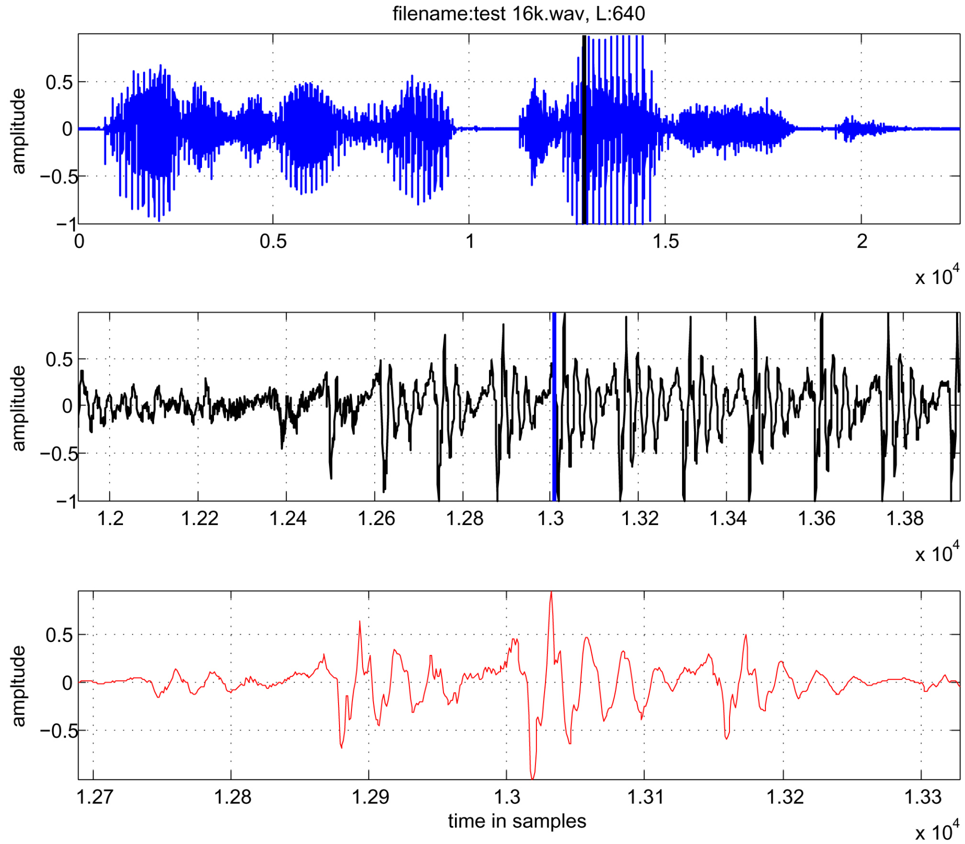

Figure 1: Sequence of waveform plots defining how the user can interactively choose a starting sample for the current analysis frame.

Cepstral Liftering – Scripted Run

A scripted run of the program 'cepstral_liftering_GUI25.m' is as follows:

1. run the program 'cepstral_liftering_GUI25.m' from the directory 'matlab_gui\cepstral_liftering',
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 selected speech file,
5. using the editable buttons, choose initial value of 40 msec for the frame length, L_m; 10 msec for the frame shift, R_m; 2048 for the value of FFT size, `nfft`;, and 30 msec for the length of the cepstrum that is preserved after low quefrency liftering, `ceplm`,
6. using the popupmenu button, choose Hamming for the short-time analysis window,
7. 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 12820 so as to match the plotted results for this example exercise,

8. hit the 'Run Current Frame' button to initiate single frame cepstral analysis of the speech beginning at the current frame starting sample, `ss`; the results of cepstral liftering analysis are shown in the various graphical plots; the top graphics panel is used to display the cepstral analysis results using the current speech frame; the middle graphics panel shows the real cepstrum of the current speech frame; the bottom graphics panel shows the STFT log magnitude spectrum along with the log magnitude spectrum that results from using a set of low quefrency lifters on the cepstrum with cutoffs varying from 20 to 100 quefrencies; 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 cepstral 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, window type, FFT size and cepstral sequence length, `ceplm`,

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_lrr.wav' is shown in Figure 2. The graphics panels show the current speech frame (top graphics panel), the real cepstrum (middle graphics panel, with the zeroth quefrency cepstral value set to 0), and the original STFT log magnitude response along with the results of cepstral liftering using a range of lifter cutoff quefrencies on the cepstrally smoothed log magnitude response (bottom graphics panel). It can be seen that over the range of lifter durations, from 20 to 100 samples in the figure, there is only a small change in the fit to the STFT log magnitude spectrum. Thus it can be seen that even a small lifter duration is adequate for spectral estimation of the vocal tract log magnitude frequency response.

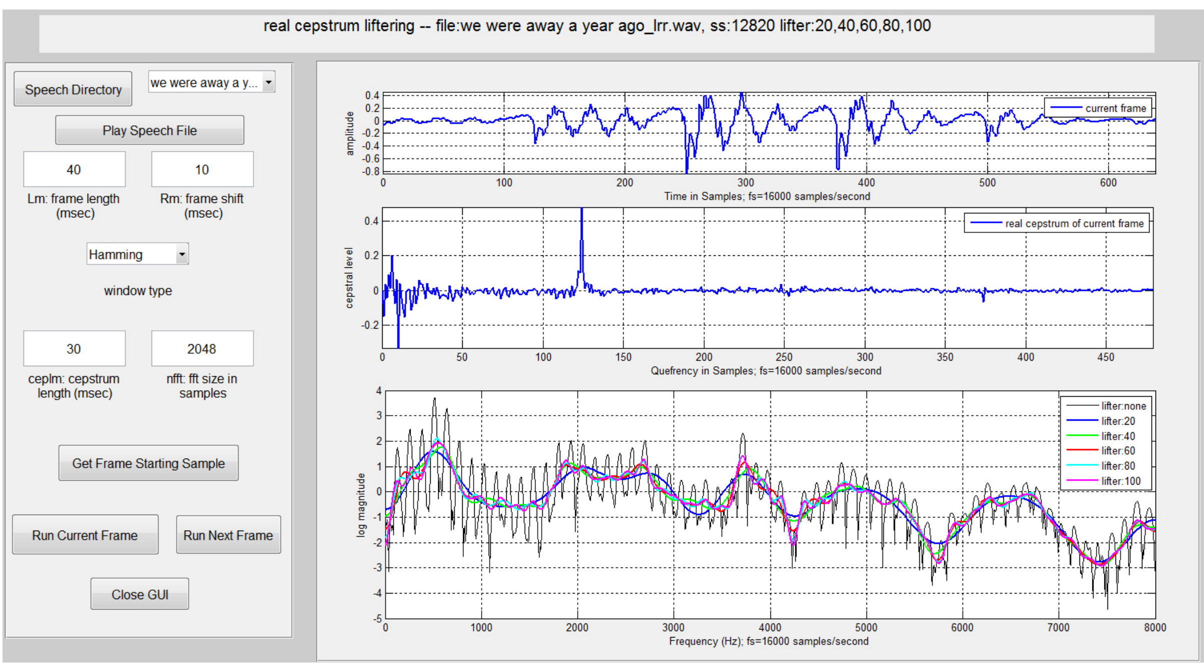

Figure 2: Example showing the effects of the cutoff quefrency of a low quefrency lifter on the smoothed log magnitude response of a frame of voiced speech. The upper graphics panel shows the current speech frame, the middle graphics panel shows the real cepstrum (with the zeroth quefrency value set to 0), and the bottom graphics panel shows the log magnitude response of the STFT of the current frame of speech, along with a series of curves of the low quefrency liftering of the cepstrum and its impact on the log magnitude frequency response.

Cepstral Liftering – Issues for Experimentation

1. run the scripted exercise above, and answer the following:

    - what is the approximate pitch period (in both samples and msec) for this frame of speech? (Hint: use the waveform plot in the top graphics panels to estimate the pitch period for the current frame of speech).
    - what is the pitch frequency (in Hz)? How is the pitch frequency manifested in the bottom spectral log magnitude plot?
    - by examining the impact of cepstral liftering with different length lifters used to spectrally smooth the log magnitude response, what can you say about the amount of spectral smoothing for lifter sizes of 20, 40, 60, 80 and 100 quefrencies?
    - how large a value of cepstral lifter duration is required for a cepstrally smoothed spectrum that captures the resonance structure of a wideband spectral analysis?
    - hold the cepstral analysis parameters to a set of fixed values and sequentially move from one analysis frame to the next frame; observe the difference in smoothed log magnitude spectrums when the frames make a transition from voiced to unvoiced segments of speech.
    - change the fft size and observe the effect on the cepstrum; can you explain what is happening as the fft size is increased, or decreased, from the nominal value of 2048 samples.

2. choose another section of speech (either the next frame which can be accessed by hitting the button ”Run Next Frame, or by hitting the button ’Get Frame Starting Sample’, and choosing a different region of speech) and repeating the exercise questions above; how do the cepstral smoothing levels for the two chosen sections of speech compare in terms of capturing the resonance structure?