MATLAB Exercise – Narrowband/Wideband Spectra of a Speech Frame

Program Directory: `matlab_gui\NB_WB_spectra`
Program Name: `NB_WB_spectra_GUI25.m`
GUI data file: `NB_WB_spectra.mat`
Callbacks file: `Callbacks_NB_WB_spectra_GUI25.m`
TADSC: Section 7.3.3, pp. 296-303

This MATLAB exercise compares and contrasts the effects of using either of two different windows (Hamming and rectangular), with two different frame durations (Narrowband–NB and Wideband–WB), on the computation of the short-time Fourier transform (STFT) of a user-selected frame of speech.

Narrowband/Wideband Spectra – Theory of Operation

This MATLAB exercise processes a user-specified frame of a speech file, with the goal of comparing linear and log magnitude spectra using either a Hamming or a rectangular window, and with a user specified set of two frame sizes-namely frame sizes suitable for wideband (10 msec) and narrowband (40 msec) analysis.

Narrowband/Wideband Spectra – 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 signal processing analysis, and for running the program.

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

1. the speech waveform on a variety of scales and resolutions,
2. the frame of speech selected by the user for analysis, along with the wideband and narrowband analysis windows,
3. the linear magnitude spectral displays of the narrowband and wideband analyses, or the log magnitude spectral displays of the narrowband and wideband analyses.

The title box displays the information about the selected file for analysis of short-time features. 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 allows the user to select either a Hamming window or a rectangular window for the short-time analysis; the default is a Hamming window,
5. an editable button that displays the user-selected starting sample (initially set to 1) for the short-time analysis (the selected sample is uses as the middle sample of the frames for both wideband and narrowband analysis),
6. an editable button that specifies the wideband frame duration, LWB, (in msec) for short-time analysis; (the default value is LWB=10 msec),
7. an editable button that specifies the narrowband frame duration, LNB, (in msec) for short-time analysis; (the default value is LNB=40 msec),

8. an editable button that specifies the wideband frame shift, RWB, (in msec) for short-time analysis; (the default value is RWB=2.5 msec),

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

10. a popupmenu button that allows the user to select the spectral display to be either log magnitude or linear magnitude response; the default is log magnitude spectral displays,

11. a pushbutton to determine the single frame starting sample, `ss`, using the iterative method described below; this starting sample (which is actually the middle sample of both wideband and narrowband spectral analyses), 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 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.

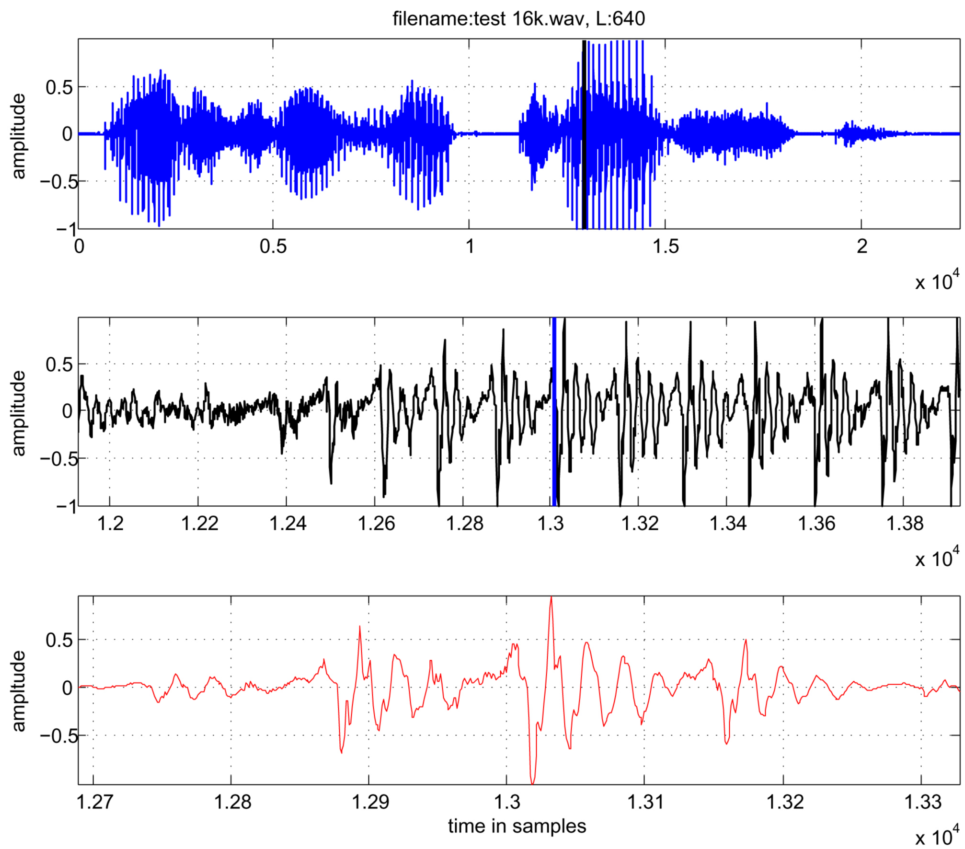

Figure 1: Sequence of waveform plots defining how the user can interactively choose a starting sample for the current analysis frame.

Narrowband/Wideband Spectra – Scripted Run

A scripted run of the program 'NB_WB_spectra_GUI25.m' is as follows:

1. run the program 'NB_WB_spectra_GUI25.m' from the directory 'matlab_gui\NB_WB_spectra',
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 analysis; choose the file 'we were away a year ago_suzanne.wav' for this example,
4. hit the pushbutton 'Play Speech File' to play the current file,
5. using the popupmenu button, select a Hamming window as the window type for short-time analysis,
6. note that the button displaying the frame middle sample, `ss`, is set initially to 1,
7. using the editable buttons, choose values of 10 msec for the wideband analysis frame length, LWB; 40 msec for the narrowband analysis frame length, LNB; 2.5 msec for the wideband analysis frame shift, RWB, and 10 msec for the narrowband analysis frame shift, RNB,
8. using the popupmenu button, select the spectrum display as log magnitude (the default),

9. 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 6319 so as to match the plotted results for this example exercise; for practical considerations the user actually chooses the initial frame middle sample for both wideband and narrowband analysis,

10. 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 WB/NB 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; a red vertical line indicates the estimate of pitch period for the current frame of speech,

11. 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+RNB`, where `RNB` is the narrowband analysis frame shift in samples,

12. experiment with different choices of speech file, and with different values for LWB, LNB, RWB, RNB, spectrum display type and window type,

13. 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 top graphics panel shows the analysis frames for wideband and narrowband analysis, and the lower graphics panel shows the log magnitude spectrums of the wideband and narrowband analyses.

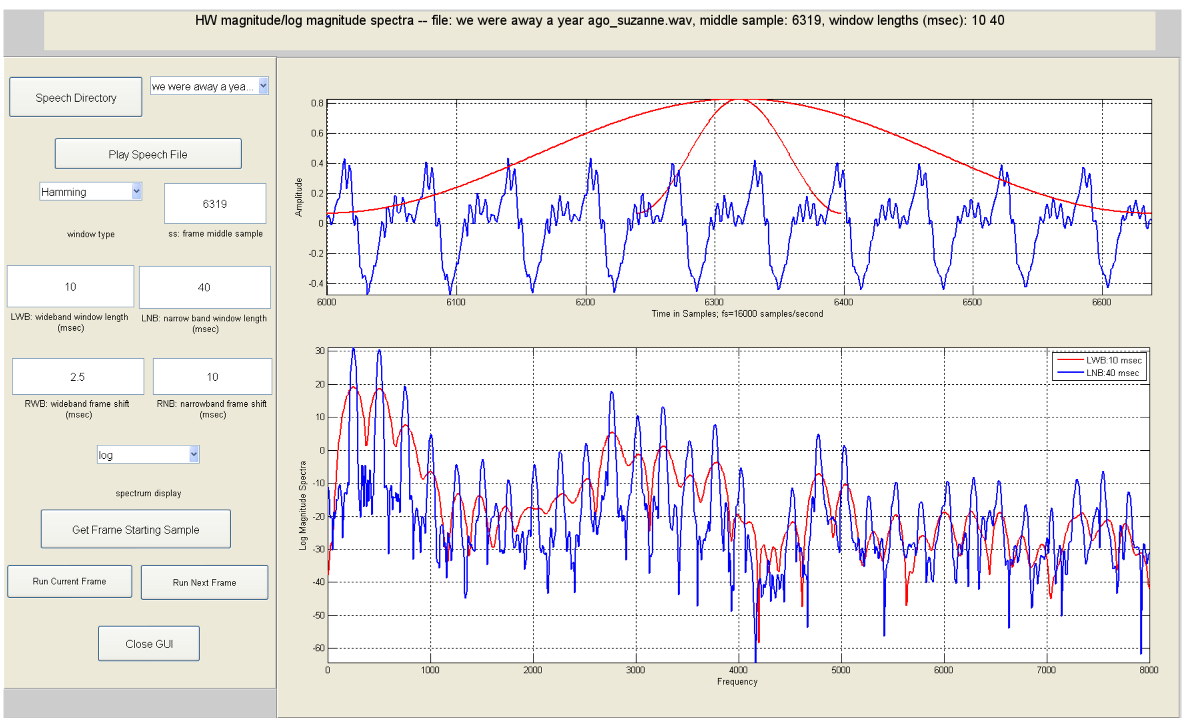

Figure 2: Short-time analysis of the speech file 'we were away a year ago_suzanne.wav'. The upper graphics panel shows the frames of speech used for wideband and narrowband spectral analysis, the lower graphics panel shows the log magnitude spectrums of the wideband and narrowband analysis,

Narrowband/Wideband Spectra – Issues for Experimentation

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

    - which window on the second graphics panel corresponds to a wideband analysis? which corresponds to a narrowband analysis?
    - what property of the selected window determines whether the window is wideband or narrowband analysis?
    - what property in the frequency domain is crucial for wideband analysis? What property in the frequency domain is crucial for narrowband analysis?
    - which display of the frame of speech in the frequency domain (i.e., the use of a linear or log magnitude scale) shows the clearest difference between narrowband and wideband spectral analysis?

2. experiment with different duration windows for the wideband and for the narrowband spectral analysis parameters

    - how small a window can be used and still provide a reasonable wideband spectral analysis of the selected frame of speech?
    - how large a window can be used and still provide a reasonable narrowband spectral analysis of the selected frame of speech?

3. change the analysis window from Hamming to Rectangular and repeat the scripted exercise

    - what is the major discernible difference in the resulting spectral analysis between using the Hamming window and using the Rectangular window?
    - would another window with less spectral leakage always provide a less noisy spectral analysis for both wideband and narrowband analyses?

    - select a speech file with a male speaker, and then select a segment of voiced speechwith the cursor. Run the spectral analysis with the default parameters.
        (a) how many pitch periods are within the wideband and narrowband windows? Can you explain the effect that this has on the corresponding spectra?
        (b) change the window lengths so that there is one pitch period in the wideband and three pitch periods in the narrowband window. Have the spectral plots changed appreciably? Explain what it happening here.
        (c) what accounts for the ”fine structure” that is evident in the narrowband spectrum?
        (d) in the narrowband spectrum, what accounts for the shape of the ”spectral lines”; i.e., the shape of the spectrum around each multiple of the fundamental frequency?
    - select a segment of unvoiced speech and repeat the analyses above.
        (a) in the narrowbnad spectrum of unvoiced speech, what accounts for the seemingly random variation (the fine structure) of the spectrum?