MATLAB Exercise – Spectrogram Plots

Program Directory: `matlab_gui\spectrogram`
Program Name: `Spectrogram_GUI25.m`
GUI data file: `Spectrogram.mat`
Callbacks file: `Callbacks_Spectrogram_GUI25.m`
TADSP: Section 7.4, pp. 312-319, Problem 7.26

This MATLAB exercise plots wideband and narrowband speech spectrograms for a user-designated speech file.

Spectrogram Plots – Theory of Operation

The sound spectrogram is one of the most fundamental tools of digital speech processing. The sound spectrogram of a speech file is an image map of the sequence of short-time log (or linear) spectrums, where each spectrum is obtained from an STFT analysis of a frame of speech, and subsequent spectrums are obtained from STFT analyses of subsequent, highly overlapped in time, frames of speech. Thus the sound spectrogram is a two-dimensional (image) plot of speech spectral information (linear or log magnitude) over the duration of the speech signal.

Spectrogram Plots – GUI Design

The GUI for this exercise consists of two panels, 3 graphics panels, 1 title box and 21 buttons. The functionality of the two panels is:

1. one panel for the graphics display,
2. one panel for parameters related to sound spectrogram parameters, and for running the program.

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

1. the speech waveform on a normalized amplitude scale,
2. a narrowband speech spectrogram of the designated speech file,
3. a wideband spectrogram of the designated speech file.

The title box displays the information about the selected file for spectrographic analysis. The functionality of the 21 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 selected file,
4. an editable button that specifies `nsec`, the duration of recording (whenever a new recording is made),
5. an editable button that specifies f_s, the sampling rate of a new recording (whenever a new recording is made),
6. an editable button with the name of an output file that will be used to save a newly recorded speech file (if that option is chosen),
7. a pushbutton to save the most recent recorded file in the chosen output file,

8. a pushbutton to begin the recording process using the user-selected parameters; when the recording button is pressed, a message box appears asking the user to click the OK button on the message box and begin recording immediately thereon; following recording, the waveform is highpass filtered and then displayed in the top graphics panel, along with a pair of graphics cursors which enable the user to select starting and ending samples for spectrographic analysis; a plot of the selected waveform is displayed and played out to the user; the user can save the file in any desired direcotry by hitting the 'Save Recording' button,

9. an editable button that specifies the wideband analysis frame duration, `LWB`, (in msec) for short-time analysis; (the default value is `LWB=3` msec),

10. an editable button that specifies the narrowband analysis frame duration, `LNB`, (in msec) for short-time analysis; (the default value is `LNB=30` msec),

11. an editable button that specifies the wideband analysis FFT size, `WB_FFT`; (the default value is `WB_FFT=1024` samples),

12. an editable button that specifies the narrowband analysis FFT size, `NB_FFT`; (the default value is `NB_FFT=1024` samples),

13. an editable button that specifies the percentage overlap between adjacent analysis frames; (the default value is `overlap=90%`),

14. an editable button that specifies the sampling frequency of the spectrogram analysis; (the default value is $f_s =$ 10000 samples per second);

15. an editable button that specifies the dynamic range (in dB) of the spectrogram display when used in the log magnitude display mode; (the default value is a dynamic range of 60 dB),

16. an editable button that specifies the scaling factor of the spectrogram display when used in the linear magnitude display mode; (the default value is a scale factor of 2),

17. a popupmenu list that lets the user choose among the two display modes, namely log magnitude and linear magnitude; (the default value is log magnitude display mode),

18. a popupmenu list that lets the user choose the STFT window type (Hamming, rectangular or triangular); (the default value is Hamming window),

19. a popupmenu list that lets the user choose either a Gray scale or a Color scale for the spectrogram display; (the default is Gray scale display),

20. a pushbutton to run the code and display the waveform and its narrowband and wideband spectrograms on the graphics panels,

21. a pushbutton to close the GUI.

Spectrogram Plots – Scripted Run

A scripted run of the program 'Spectrogram_GUI25.m' is as follows:

1. run the program 'Spectrogram_GUI25.m' from the directory 'matlab_gui\spectrogram',

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 'Play' button to play out the selected speech file;

5. if you want to record a new speech file for spectrogram analysis, use the set of 5 buttons to specify the number of seconds of recording (4 is the default), the recording sampling rate (10000 is the default), the file in which to save the recording ('out_file' is the default), and then hit the 'Record' button (to create a new speech recording), and then hit the 'Save Recording' button to save the recording in the designated output file,

6. using the editable buttons, use the default values for the wideband and narrowband spectrogram parameters including wideband window length, `L_WB` (3 msec), narrowband window length, `L_NB` (30 msec), wideband FFT length (1024 samples), narrowband FFT length (1024 samples), percentage window overlap (90%), spectrogram sampling rate (10000 Hz), log magnitude spectrum dynamic range (60 dB), scale factor for linear magnitude dynamic range (2), log magnitude display, Hamming window type, gray scale spectrogram;

7. hit the 'Create Spectrogram' button to compute the narrowband and wideband spectrograms and display them (along with the speech waveform) on the graphics panels,

8. experiment with different choice of speech file, and with different values for all the spectrogram parameters,

9. hit the 'Record Button' to record up to nsec=4 seconds of speech; using the graphics cursors to endpoint the recorded utterance, the endpointed waveform is plotted in the top graphics panel; the endpointed speech array can be saved in any directory and with any desired filename by setting the filename in the output file box, and by hitting the 'Save Recording' button,

10. 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. The graphics panels show the speech waveform (top graphics panel), the narrowband spectrogram (middle graphics panel), and the wideband spectrogram (bottom graphics panel).

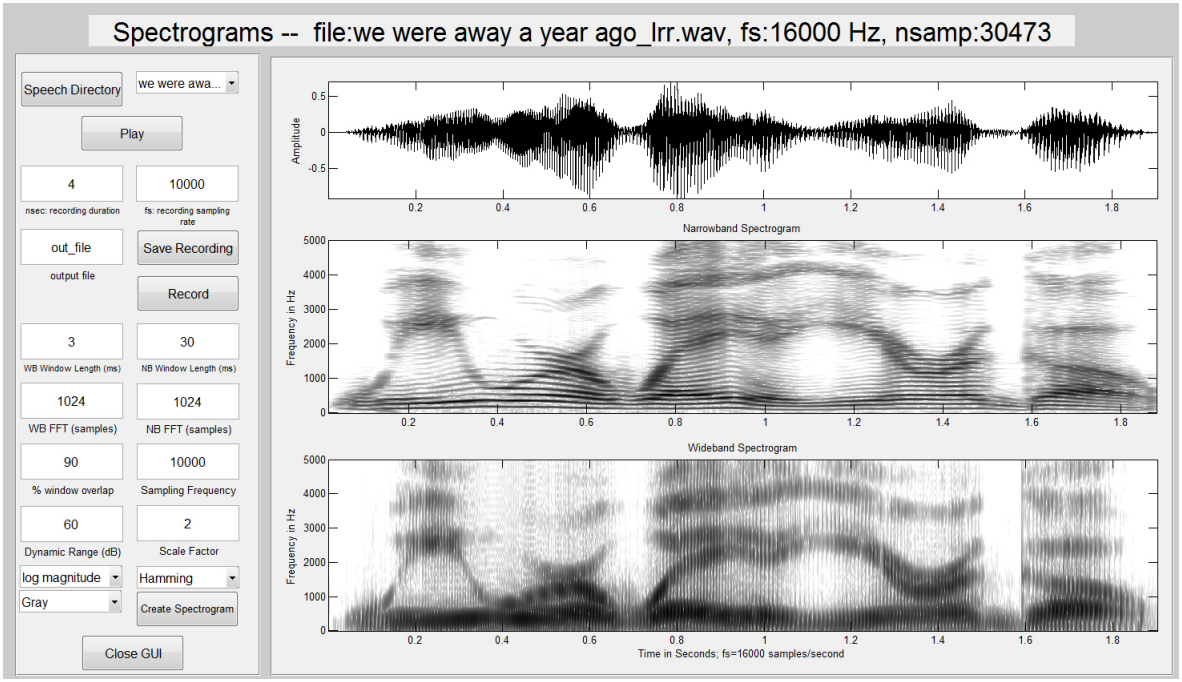

Figure 1: Graphical output obtained from this exercise using the speech file 'we were away a year ago_lrr.wav' is shown in Figure 1. The graphics panels show the speech waveform (top graphics panel), the narrowband spectrogram (middle graphics panel), and the wideband spectrogram (bottom graphics panel).

Spectrogram Plots – Issues for Experimentation

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

   - what is the key property of the window that determines whether the spectral analysis is narrowband or wideband?
   - how does this key property of the window vary with respect to the average pitch period of the speaker for whom the spectrogram is computed?
   - what happens to the spectrogram displays when the dynamic range of the graphics plot is varied by $\pm 10 - 20$ dB?
   - which form of spectrogram (i.e., gray scale or color) appears to be more information bearing; i.e., which best shows off the formants (for wideband analysis) and which best shows off the pitch harmonics (for narrowband analysis)?
   - what is the impact of switching a rectangular window for the Hamming window? What is the impact of switching a triangular window for the Hamming window?
   - what happens when the sampling rate of the processing is changed from the default value of $f_s = 10,000$ samples per second, to a sampling rate of $f_s = 16,000$ samples per second?

2. using the speech file 's5_edited.wav', rerun the exercise and create new spectrograms

   - do the spectrogram plots appear to be as information bearing as for the previous speech file?
   - what causes the spectrogram plots to lose clarity of spectral representation, for both narrowband and wideband spectrograms?
   - how would you adjust the frame durations for narrowband and wideband spectral analysis for this utterance?

3. run the exercise with varying the window overlap parameter from the default value of 90% overlap, to an overlap of 50%;

   - what phenomenon occurs as the percentage window overlap falls from 90% to 80% and then to

1. select the speech file 's5_edited.wav' and run the spectrogram program with the default parameters.
2. can you identify the voiced regions by looking at the wideband spectrogram? What distinguishes voiced speech regions from unvoiced speech regions?
3. in the voiced regions of the narrowband spectrogram, what is the vertical spacing between the lines that move almost horizontally?
4. change the % window overlap parameter from 90 to 50. What effect does this have on the spectrograms¿
5. change the narrowband window length to the same value as the length of the wideband spectrogram, and run the program again. What do you observe? Is there any difference between the two spectrograms? What do you conclude is the only difference between the narrowband and wideband spectrograms?
6. how should the fft length be chosen relative to the window length? What are the issues in setting the fft length? Does the fft length affect the frequency resolution?
7. which spectrogram shows events in time better, the wideband or narrowband spectrogram?