1.5 MATLAB Exercise – Speech Recording, Playback, Plotting

Program Directory: `matlab_gui\record_display_speech`
Program Name: `record_display_speech_GUI25.m`
GUI data file: `record_display_speech.mat`
Callbacks file: `Callbacks_record_display_speech_GUI25.m`

This section describes a MATLAB exercise that uses the file read command to read in an existing speech file, the record function to record a speech signal, and the file save command to save the results in a designated file on any desired MATLAB directory. The program plots the (existing or newly recorded) speech signal and its short-time log energy and zero crossings rate contours. Finally, for newly recorded signals, the MATLAB program uses a graphical cursor to manually segment the recorded speech signal, as desired; e.g., eliminate background signal both before and after the desired speech signal, prior to saving the recorded speech in the designated directory and in the designated file.

Speech Recording, Playback, Plotting – GUI Design

The GUI for this exercise consists of four panels, three graphics panels, 1 title box, and 11 buttons. The functionality of the four panels is:

1. one panel for the graphics display,
2. one panel for parameters related to retrieval, display and plotting for an existing speech file,
3. one panel for recording, plotting and saving newly recorded speech that is manually endpointed and saved in a designated output file in a user designated directory,
4. and one panel for closing the GUI.

The functionality of the three graphics panels is:

1. one graphics panel to display the waveform from either an existing speech file or from a newly recorded speech waveform,
2. one graphics panel to display the short-time log energy contour and the short-time zero crossings contour of the current speech signal,
3. and one graphics panel to display a wideband spectrogram of the current speech signal.

The title box displays information about the file used in this exercise. Finally the functionality of the set of 11 buttons is:

1. a pushbutton to select the directory where a large set of speech files are stored,
2. a popupmenu button to display the set of available speech files in the designated speech files directory. The user scrolls through all the entries of the popupmenu list and selects one file for analysis and display,
3. a pushbutton to plot the selected speech file samples and display the short-time log energy and zero crossings contours of the speech signal,
4. a pushbutton to play the current speech file,
5. an editable text button to specify the sampling rate, f_s, of a new recording (initially pre-set to $f_s = 16,000$ Hz),
6. an editable text button to specify the number of seconds of recording (initially pre-set to `nsec=3` seconds),
7. a pushbutton to begin the recording process,

8. a pushbutton to begin the process of endpointing the recorded speech file,

9. an editable text button with the name of the file in which the newly recorded speech is optionally stored (in the user designated directory); the filename is initially set to 'out_record',

10. a pushbutton to initiate the process of saving the recorded speech file to the designated directory using the designated filename,

11. a pushbutton to terminate the exercise (closing the GUI)

Speech Recording, Playback, Plotting – Scripted Run

A scripted run of the program 'record_display_speech_GUI25.m' is as follows:

1. run the program 'record_display_speech_GUI25.m' from the local directory 'matlab_gui\record_display_speech',

2. hit the pushbutton labeled 'Speech Directory' and click on the directory 'speech_files' (the file system display should be pointing to this default directory, called 'speech_files', where most speech files are stored),

3. select the speech file 's5_edited.wav' from scrolling through the list of files in the popupmenu button,

4. hit the 'Plot Speech' button to plot the speech waveform in the upper graphics panel, and the speech signal short-time log energy and zero crossings contours in the middle graphics panel, and a wideband spectrogram of the speech signal in the bottom graphics panel,

5. hit the 'Play Speech' button to play and listen to the current speech file,

6. edit any or all of the three editable buttons prior to recording a new speech file; i.e., to change the recording sampling rate, the number of seconds of recording, or the name of the output speech file in which the recorded signal will be saved, and then hit the 'Record Speech' button. A message box will appear on the screen telling the user to hit the OK part of the message in order to begin recording,

7. hit the 'Endpoint Speech' button to allow the user to interact with a display of the speech waveform, via a pair of graphic cursors, in order to edit the recorded speech file and remove undesirable sections of background signal; following the editing process, a graphics plot of the recorded waveform and its short-time log energy and zero crossings rate contours will appear in the upper pair of graphics panels, and a wideband spectrogram of the speech file will appear in the lower graphics panel; this editing process can be repeated as often as desired using the steps outlined above,

8. if the user wants to save the manually segmented speech waveform in a user chosen directory and the named speech output file, the user hits the pushbutton 'Save Speech'. The speech file is saved in a user designated directory of MATLAB which is selected via the `uigetdir` MATLAB command. If the recorded speech is not adequate for the user's needs, the recording process can be iterated until an acceptable recording is obtained,

9. the steps outlined above can be iterated on any previously recorded file or on any newly recorded file,

10. hit the 'Close GUI' button to terminate the run.

The resulting graphical display for the speech file 's5_edited.wav' is shown in Figure 1.

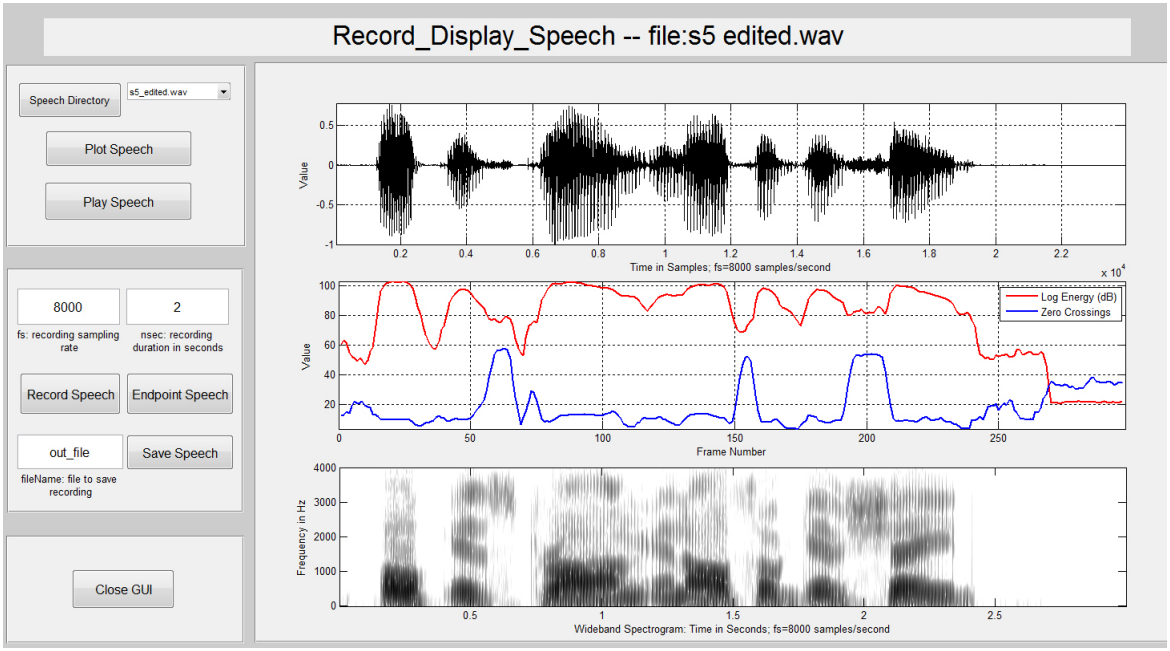

Figure 1: Example of graphical output of program 'record_display_speech' for the speech file 's5_edited.wav'

Speech Recording, Playback, Plotting – Issues for Experimentation

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

   - what was the text of the speech utterance in the file `'s5_edited.wav'`?
   - what are (approximately) the beginning and ending samples of the speech utterance?
   - the sampling rate of the selected file is known to be $f_s = 8000$ samples per second; using this sampling rate what is the duration of the speech utterance (in seconds) and of the total sound within the file?
   - if the peak log energy (in dB) of the signal within the file is scaled to 0 dB, what is the average log energy in the background regions (before and after the speech utterance), and at the end of the file?
   - can you identify the sounds in the utterance for which the short-time zero crossing rate exceeded a threshold of 30?

2. attach a microphone (or use the built in microphone of your computer, when appropriate), and record the utterance `'We were away'` at the default sampling rate of $f_s = 16,000$ samples per second.

   - using the graphics cursor, set the initial and final speech utterance frames and save the recorded and endpointed speech in the file `'out_record'`. (Note that the output file is saved in a user designated directory.) What is the duration of the resulting speech utterance, in seconds?
   - how often does the short-time zero crossings parameter exceed the threshold of 30? Why is this the case?
   - how much does the short-time log energy plot change over the duration of the speech utterance? Why is this the case?

3. change the sampling rate from $f_s = 16,000$ samples per second to $f_s = 6000$ samples per second, and re-record the utterance `'We were away'`.

   - can you detect the difference in sound quality between the speech at these two different sampling rates? What is the main effect that you hear that enables you to tell the difference in sound quality?

- record the utterance ”657 0892” at sampling rates of 6 kHz, 8 kHz, 10 kHz and 16 kHz. Edit the recorded signal to remove all initial and final background signal and just preserve the region of the speech signal. Save all the signals (the 4 versions of the recorded utterance) in the MATLAB directory of your choice.
- using the playback mode capabilities of this exercise, load the speech files from the current directory and listen to all of the recordings of the recorded utterance in the previous step and describe how the signal quality changes with changes in sampling rate.