MATLAB Exercise – Time Domain Features

Program Directory: `matlab_gui\time_domain_features`
Program Name: `time_domain_features_GUI25.m`
GUI data file: `time_domain_features.mat`
Callbacks file: `Callbacks_time_domain_features_GUI25.m`
TADSC: Sections 6.3-6.4, pp. 248-265, Problem 6.17

This MATLAB exercise examines the behavior of a set of three time domain, short-time features, as a function of the type of analysis window, the frame duration and the frame shift. The short-time features used in this exercise are short-time linear/log energy, short-time linear/log magnitude, and short-time zero crossing rate (per 10 msec interval).

Time Domain Features – Theory of Operation

This MATLAB exercise performs short-time analysis on a designated speech file using frames of specified duration, analysis window type, and frame shift. The features that are measured are short-time linear/log energy, short-time linear/log magnitude, and short-time zero crossing rate per 10 msec interval. The exercise runs the time domain feature code for a user-specified number of times, (up to a maximum of 5 times), with a user-specified initial frame duration for the first iteration, and with doubling the frame duration for each subsequent iteration. Thus if we want four runs, the first analysis run begins with a frame duration of L msec, and a frame shift of R msec; the second run uses a frame duration of $2*L$ msec and a frame shift of R msec; the third run uses a frame duration of $4*L$ msec and a frame shift of R msec; and finally the fourth run uses a frame duration of $8*L$ msec and a frame shift of R msec. The results of each iteration are plotted on a common graphics panel, thereby enabling the user to compare and contrast the impact of varying frame duration on the resulting short-time feature sets.

The MATLAB exercise displays the following signals:

Graphics Panel1 : the normalized speech waveform

Graphics Panel2 : the short-time linear or log energy contour for each of the varying frame durations,

Graphics Panel3 : the short-time linear or log magnitude contour for each of the varying frame durations,

Graphics Panel4 : the short-time zero crossing rate per 10 msec interval for each of the varying frame durations.

Time Domain Features – GUI Design

The GUI for this exercise consists of two panels, 4 graphics panels, 1 title box and 11 buttons. The functionality of the two panels is:

1. one panel for the graphics display,
2. one panel for parameters related to the set of time domain features, and for running the program.

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

1. the speech waveform on a normalized amplitude scale,
2. a set of up to five plots of the linear or log short-time energy (on a dB scale for the log energy plots) for the up to five different frame lengths,
3. a set of up to five plots of the linear or log short-time magnitude (on a dB scale for the log magntiude plots) for the up to five different frame lengths,
4. a set of up to five plots of the zero crossings per 10 msec interval rate for the up to five different frame lengths.

The title box displays the information about the selected file for analysis of short-time features. The functionality of the 11 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 push button that allows the user to play the selectedspeech file,
4. an editable button that specifies the initial frame duration, L, (in msec) for short-time analysis; (the default initial frame duration is $L = 5$ msec),
5. an editable button that specifies the frame shirt, R, (in msec) for short-time analysis; (the default value is $R = 10$ msec),
6. an editable button that specifies the number of doublings of the short-time analysis (up to a maximum of 5 doublings), where each run doubles the duration of the frame from the previous value; the default is four doublings of L,
7. an editable button that specifies the sampling rate, f_s, of the selected speech file,
8. a popupmenu button that lets the user choose either linear or log magnitude frequency response; (default is log scaling),
9. a popupmenu button that lets the user choose either a Hamming or rectangular window as the short-time analysis window; (default is Hamming window),
10. a pushbutton to run the code and display the results of short-time analysis on the four graphics panel displays,
11. a pushbutton to close the GUI.

Time Domain Features – Scripted Run

A scripted run of the program 'time_domain_features_GUI25.m' is as follows:

1. run the program 'time_domain_features.m' from the directory 'matlab_gui\time_domain_features_gui25',
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 'test_16k.wav' for this example,
4. hit the pushbutton 'Play Speech File' to play out the current file,
5. using the editable buttons, choose an initial value of 5 msec for the frame length, L; choose an initial value of 10 msec for the frame shift, R, and an initial value of 4 for the number of doublings of frame length for the complete run for the number of runs where the frame length doubles each subsequent run,
6. using the popupmenu button, choose log scaling of the short-time energy and short-time magnitude contours; similarly using the other popupmenu button, choose Hamming for the short-time analysis window,
7. hit the 'Run' button to compute and display the contours of short-time energy, short-time magnitude and short-time zero crossing rate per 10 msec interval on the appropriate graphics panels,
8. experiment with different choices of speech file, and with different values for L, R, window type, number of runs, and scaling option,
9. hit the 'Close GUI' button to terminate the run.

An example of the graphical output obtained from this exercise using the speech file 'test_16k.wav' is shown in Figure 1. The graphics panels show the speech waveform (top graphics panel), the short-time log energy contour (second graphics panel), the short-time log magnitude contour (third graphics panel), and the short-time zero crossings rate per 10 msec interval (fourth graphics panel). Each of the last three panels shows the results of short-time signal analysis for the four different values of L.

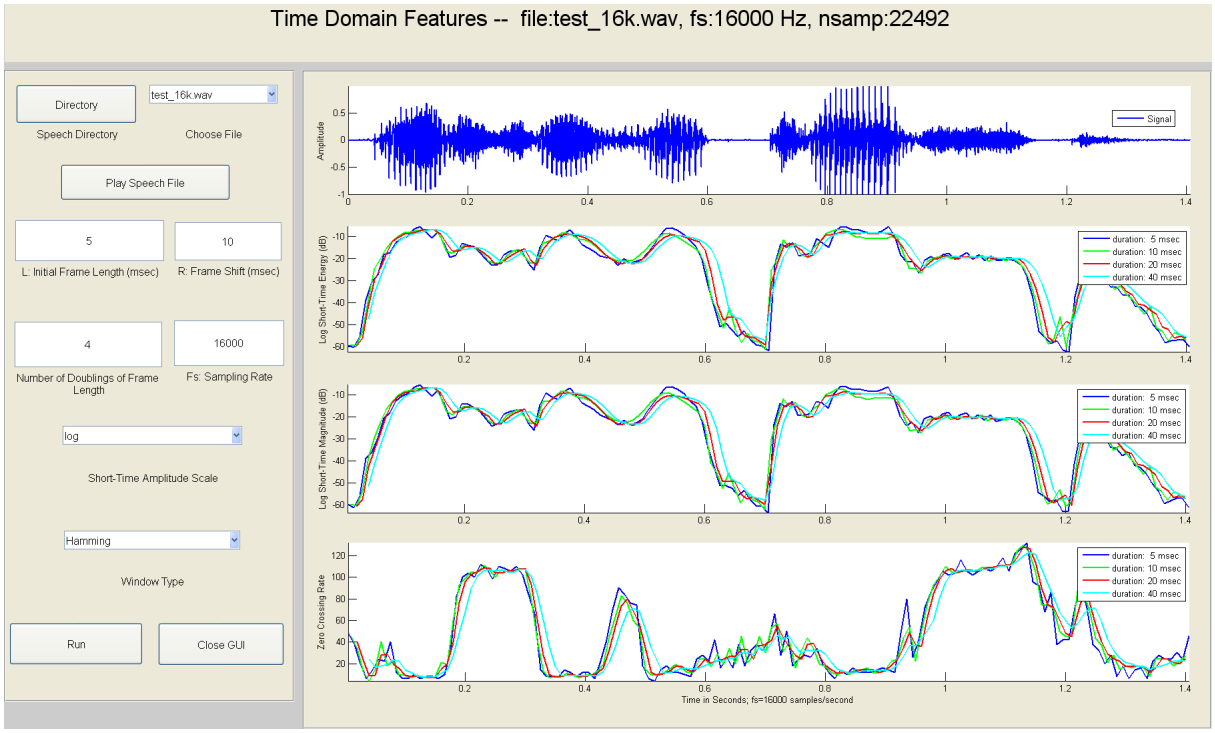

Figure 1: Short-time analysis of the speech file 'test_16k.wav' for four different values of frame duration. The graphics panels show the speech waveform (top graphics panel), the short-time log energy contour (second graphics panel), the short-time log magnitude contour (third graphics panel), and the short-time zero crossings rate per 10 msec interval (fourth graphics panel). Each of the last three panels shows the results of short-time signal analysis for the four different values of L.

Time Domain Features – Issues for Experimentation

1. select the digit file '6A.waV' and run the program with the default parameters, and the scale set to 'lin'.
2. as the window length increases, how does the appearance of the curves change? Why does this occur?
3. where is the zero crossing rate the highest? What type of speech sounds occur at these points in time?
4. where are the log energy and log magnitude contours the highest? What type of speech sounds occurs at these points in time?
5. select the speech file 's5_edited.wav'. Run the program and note the sounds in the speech utterance. Which phonetic features of speech are easily noted in these plots? Can you easily locate unvoiced speech sounds, low level speech sounds, areas of background signal?

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

- why do the curves of short-time log energy and short-time log magnitude appear so similar?
- at what duration do the short-time log energy and short-time log magnitude contours become relatively stable (loss of high frequency variability)?
- there are 3 regions where the short-time zero crossing rate is large (order of 60 or greater); what are the sounds in the speech utterance 'This is a test' that correspond to each of these three high zero crossing rate regions?

2. change the short-time amplitude scale from log to lin and rerun the exercise
   - what impact does this change in scale have on the short-time feature plots?
3. change the short-time amplitude scale back to log and change the window type to 'Rectangular' window and rerun the exercise
   - what impact does this change in window type have on the short-time feature plots?
4. change the frame shift parameter, R, from 10 msec to 2 msec and rerun the exercise
   - what impact does this change in scale have on the short-time feature plots?