MATLAB Exercise – Cepstral Pitch Detector

Program Directory: `matlab_gui\cepstral_pitch_detector`
Program Name: `cepstral_pitch_GUI25.m`
GUI data file: `cepstral_pitch.mat`
Callbacks file: `Callbacks_cepstral_pitch_GUI25.m`
TADSP: Section 10.5.6, pp. 625-632, Problem 10.9

This MATLAB exercise implements a pitch period detector based on detecting and tracking peaks in the (real) cepstrum during regions of voiced speech. The major feature of this pitch period detector is the use of a secondary cepstral peak detector, for each frame of speech, in order to detect and correct pitch period detection errors due to effects such as pitch period doubling or halving and related phenomena.

Cepstral Pitch Detector – Theory of Operation

Given a designated speech file, with designated frame duration, and frame shift, this exercise estimates (on a frame-by-frame basis) the location, `pp1`, and amplitude, `plevel1`, of the largest cepstral peak in a gender-specified range of the cepstrum. The exercise next zeros out a specified region in the vicinity of the largest cepstral peak and detects the location, `pp2`, and amplitude `plevel2`, of the second largest cepstral peak, again in a gender-specified range of the cepstrum. Using the ratio of the amplitudes of the first and second cepstral peaks as a measure of confidence in the location of the cepstral peak, the exercise classifies each first cepstral peak as either being above (`ppdf=1`) or below (`ppdf=0`) the user specified cepstral peak ratio threshold, `pthr1`. Using the 1-bit (`ppdf`) contour, 'regions' of cepstral peaks which are above threshold are detected and used to extend the regions of voiced speech to the left of the initial frame of each region, and to the right of the final frame of each region, based on whether the secondary peak in the extended range is within a specified percentage of the primary pitch period. This pitch period smoothing method generally will extend one or more pitch period regions by a few frames and is especially useful at the beginning and ending of voiced speech regions.
A median smoother is applied to both the cepstral pitch period contour and to the confidence score derived from taking the ratio of the first valid cepstral peak to the second cepstral peak, for each frame, in order to eliminate single and double frame pitch period errors and irregularities. The MATLAB program plots both the unsmoothed results (pitch period and confidence score contours) along with the median smoothed pitch period contour and the median smoothed confidence score derived from the ratio of first-to-second cepstral peak level.

Cepstral Pitch Detector – 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 the cepstral analysis, and for running the program.

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

1. the speech waveform,
2. the unsmoothed and smoothed pitch period contour of the utterance,
3. the unsmoothed and smoothed confidence score contours of the utterance.

The title box displays the information about the selected file. 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 play the current speech file,
4. an editable button that specifies the sampling rate, `fsd`, of the signal processing parts of the pitch detector; (the default value is `fsd=6000` samples per second),
5. a popupmenu button that specifies the gender of the talker; (the default is male),
6. an editable button that specifies the frame duration, L_m, (in msec) for short-time analysis; (the default value is $L_m = 40$ msec),
7. an editable button that specifies the frame shift, R_m, (in msec) for short-time analysis; (the default value is $R_m = 10$ msec),
8. an editable button that specifies the cepstral ratio threshold, `pthr1`, for high confidence that the detected cepstral pitch peak is a true pitch period for the current frame; (the default value for `pthr1` is 2.5),
9. an editable button that specifies the length of the median smoother, `Lmed`, in frames: (the default value for `Lmed` is 5 frames),
10. a pushbutton to run the code and display the speech waveform, the raw pitch period contour, and the raw confidence scores for the primary and secondary cepstral pitch period estimates, along with the ratio of primary-to-secondary cepstral scores, for each frame of speech,
11. a pushbutton to perform median smoothing on the pitch period contour and the confidence frame contour, and to plot the results (both unsmoothed and smoothed pitch period estimates, and smoothed confidence scores),
12. a pushbutton to close the GUI.

Cepstral Pitch Detector – Scripted Run

A scripted run of the program 'cepstral_pitch_GUI25.m' is as follows:

1. run the program cepstral_pitch_GUI25.m' from the directory 'matlab_gui\cepstral_pitch_detector',
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 the original speech file,
5. using the editable buttons, choose initial value of 10000 samples per second for the parameter `fsd`; an initial value of 40 msec for the analysis frame duration, L_m; an initial value of 2.5 for the pitch threshold, `pthr1`; and an initial value of 5 frames for the median smoother duration, `Lmed`,
6. using the popupmenu button specify the gender of the speaker as male in this example,
7. hit the 'Run Cepstral Pitch' button to compute and display the signal waveform in the top graphics panel; the raw (unsmoothed) pitch period contour, and the raw (unsmoothed) confidence score contour,
8. hit the 'Smooth Cepstral Pitch' button to median smooth the raw pitch period contour and the raw confidence score, and re-plot these contours,
9. experiment with different choices of speech file, and with different values for `fsd`, L_m, R_m, `pthr1` and `Lmed`,
10. 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 Figures 1- 2. The graphics panels in Figure 1 shows the speech waveform (top graphics panel), the raw (unsmoothed) pitch period contour (middle graphics panel, and the raw (unsmoothed) confidence score contours (bottom graphics panel) which consist of the cepstral peak level contour, the cepstral second peak level contour, the peak ratios of the two cepstral levels (divided by a factor of 25 so as to allow all three curves to be shown on a single graphics panel), and the cepstral decision threshold, `pthr1`, again divided by 25 to be consistent with the cepstral peak ratio curve).

The graphics panel in Figure 2 shows the speech waveform (top graphics panel), the median smoothed pitch period contour (middle panel), and the median smoothed confidence score (bottom panel) where the confidence score is derived from the ratio of the top and secondary cepstral peaks.

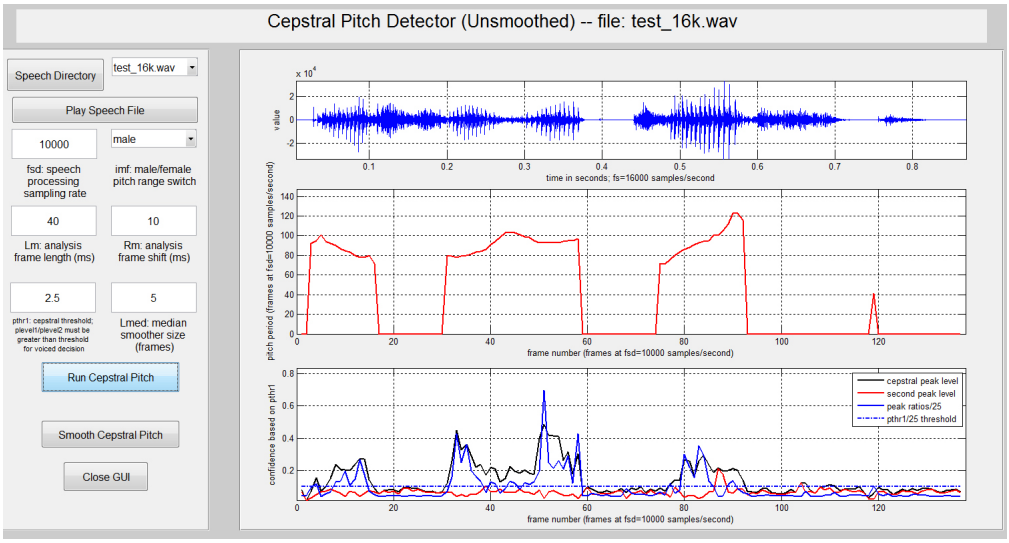

Figure 1: Graphical output from cepstral pitch detection algorithm. The graphics panels show the waveform, the unsmoothed pitch period contour and the unsmoothed confidence score contours for the utterance 'This is a test'.

0.0.1 Cepstral Pitch Detector – Issues for Experimentation

1. Compare the waveform plot (especially when using the exercise 'strips_plot') to the pitch period contour from this exercise. Does the pitch period contour seem to be shifted with respect to the waveform? If so, what could account for this phenomenom?

2. Increase or decrease the window length and observe how the estimates of pitch period vary as the window gets larger or smaller. What happens when you keep the frame length, L_m, the same, but make the frame shift, R_m, smaller and smaller?

3. What is the impact of the input parameter `pthr1` on the estimates of pitch period, especially at the boundaries between voiced and unvoiced regions of speech?

4. Experiment with the median length parameter (`Lmed`) to see the tradeoff between removing outlier pitch period estimates and over-smoothing the good pitch period estimates.

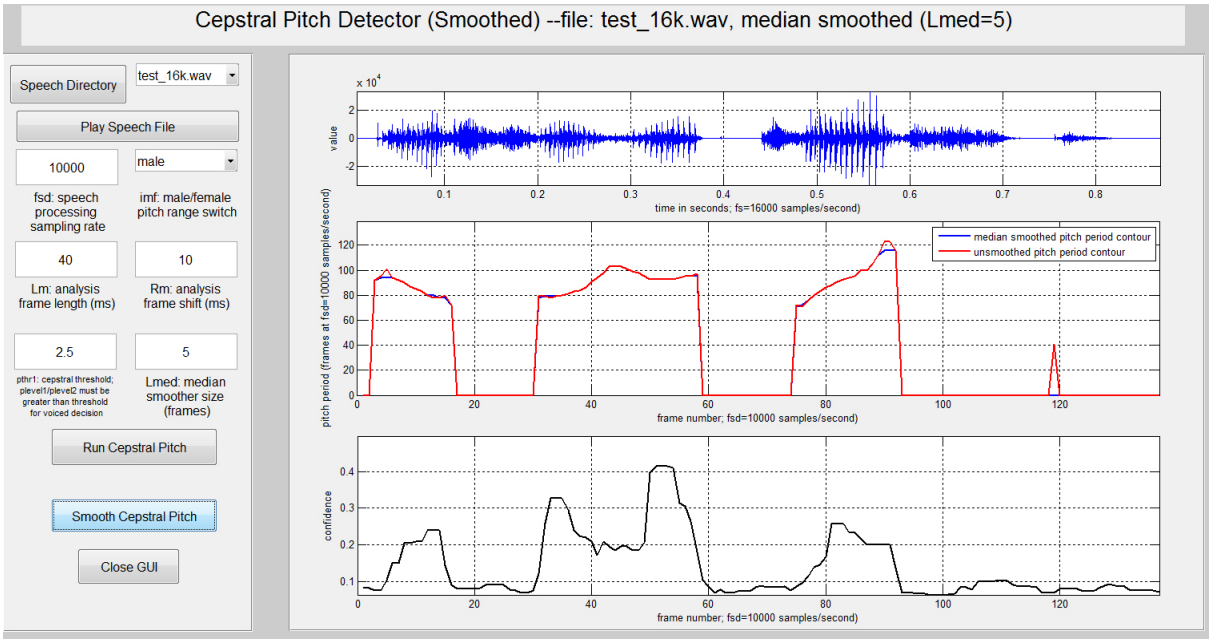

Figure 2: Graphical output from cepstral pitch detection algorithm. The graphics panels show the waveform, the median smoothed pitch period contour and the median smoothed confidence score contour for the utterance 'This is a test'.