

MATLAB Exercise – Phase Vocoder

Program Directory: matlab_gui\phase_vocoder

Program Name: phase_vocoder_GUI25.m

GUI data file: phase_vocoder.mat

Callbacks file: Callbacks_phase_vocoder_GUI25.m

TADSP: Section 7.11.3, pp. 372-379, Section 12.2.2, pp. 848-850

This MATLAB exercise implements a phase vocoder with the capability of speed-up or slow-down of a speech or audio signal by a factor, r , which varies from $r = 0.25$ (slow-down by factor of 4) to $r = 4.0$ (speed-up by a factor of 4).

Phase Vocoder – Theory of Operation

This exercise implements a classic phase vocoder with the capability of speed-up or slow-down of a speech or audio file by interpolating the Short-Time Fourier Transform (STFT) magnitude and calculating the phase advance for the signal with speed-up/slow-down rate of r , where $r < 1$ slows down the signal, and $r > 1$ speeds it up.¹

Phase Vocoder – GUI Design

The GUI for this exercise consists of two panels, two graphics panels, 1 title box and 7 buttons. The functionality of the two panels is:

1. designate a speech or audio file for speed-up or slow-down, set the analysis frame length, choose a value for the speed-up/slow-down rate, and run the exercise,
2. plot wideband spectrograms of the original speech signal and the speeded-up/slowed-down signal in the two graphics panels.

The title box displays generic information about MATLAB exercise. The functionality of the 7 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. an editable button that specifies the speed-up/slow-down factor; (default value of $r = 0.5$),
4. an editable button that specifies the analysis frame length, wlen, in msec; (default is wlen=60 msec),
5. a pushbutton to run the code and plot wideband spectrograms of the original speech signal and the speeded-up/slowed-down signal,
6. a pushbutton to listen to the original signal and the speeded-up/slowed-down signal, in sequence,
7. a pushbutton to close the GUI.

¹This program was written by Prof. Dan Ellis of Columbia University, and modified slightly by the authors of this MATLAB Exercise guide. The original reference is www.ee.columbia.edu/dpwe/resources/matlab/pvoc/ and the description of the code and the processing within the code is contained in the document "A Phase Vocoder in MATLAB".

Phase Vocoder – Scripted Run

A scripted run of the program 'phase_vocoder_GUI25.m' is as follows:

1. run the program 'phase_vocoder_GUI25.m'_rev1_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. using the editable buttons, set the initial values for the rate parameter to $r = 0.5$, and for the analysis frame length to $wlen=60$ msec,
5. hit the 'Run Phase Vocoder' button to compute the phase vocoder parameters for slowing down the utterance by a factor of $r = 2$; wideband spectrograms of the original signal and the speeded-up/slowed-down signals are given in the pair of graphics panels,
6. hit the 'Play Original/Phase Vocoder' button to play out in sequence the original speech file, followed by the speeded-up/slowed-down speech file,
7. experiment with different choices of speech file, and with different values for r and $wlen$,
8. hit the 'Close GUI' button to terminate the run.

The program runs very efficiently on virtually any speech or audio file and provides an exceptionally clean signal for the entire range of speed-up and slow-down factors.

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 wideband spectrograms of the original signal (top graphics panel) and the speeded-up signal (by a factor of 2.5 in this example) in the lower graphics panel.

Phase Vocoder – Issues for Experimentation

1. using the default parameters, run the program and play the processed speech signal. Does it sound natural? Can you hear any artifacts (non-speech-like sounds) of the processing?
2. by changing the speed-up/slow-down factor to the range $r=0.25$ to 1.0, at what level of slow-down factor does the speech distortion become noticeable? At what point is it sufficiently annoying that you would consider the speech unacceptable?
3. repeat the last step for speed-up factors of $r=1$ to $r=4$, At what value of r do you get noticeable speech signal distortion? At what value of r is the speed annoying and of unacceptable quality?
4. what can you see in the spectrogram plots by comparing wideband spectrograms for both speeded-up and slowed-down processing? Does the spectrogram of the speed changed signal retain the usual properties for voiced and unvoiced sounds?
5. repeat the above analyses on different speech files and on the audio file 'Maple_short.wav'. Is it easier or harder to find the distortion of an audio file than of a speech file?

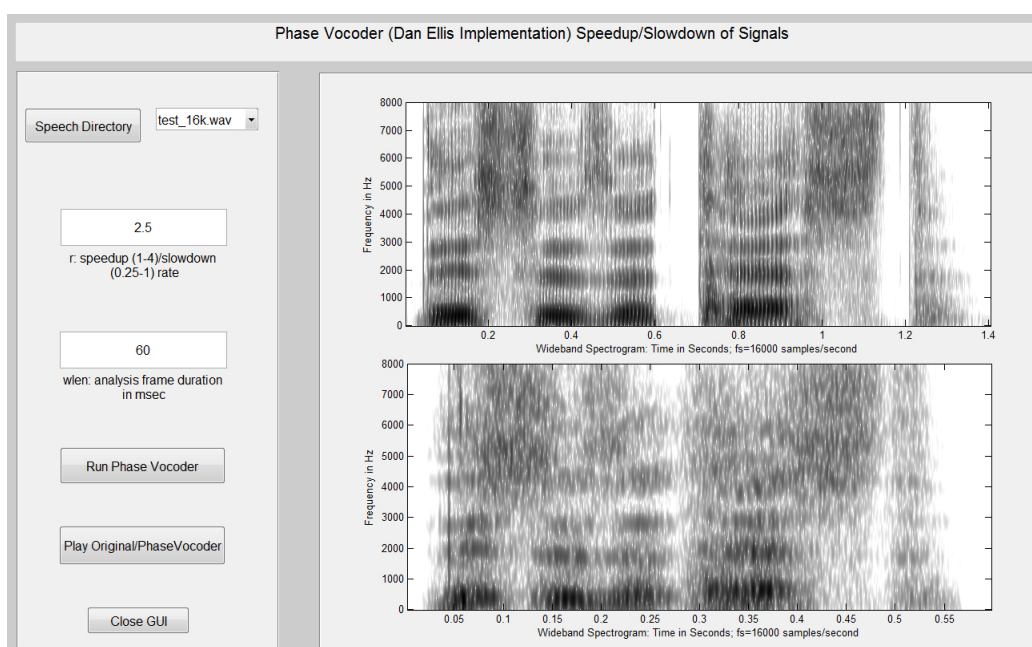


Figure 1: Graphical output from phase vocoding of a speech file that speeds up the speech by a factor of $r=2.5$. The graphics panels show wideband spectrograms of the original signal (top graphics panel) and the speeded-up signal by a factor of $r=2.5$ (lower graphics panel).