MATLAB Exercise – Impact of Spectral Magnitude and Phase on Long-Time Analysis and Synthesis of Speech

Program Directory: `matlab_gui\phase_magnitude`
Program Name: `phase_magnitude_GUI25.m`
GUI data file: `phase_magnitude.mat`
Callbacks file: `Callbacks_phase_magnitude_GUI25.m`

This MATLAB Exercise illustrates the relative importance of long-time signal phase and magnitude on the intelligibility of a speech signal.

Spectral Magnitude/Phase – Theory of Operation

It is often tacitly assumed that the role of long-time spectral magnitude is significantly greater than the role of long-time spectral phase when performing a long-time analysis of a speech signal, i.e., it is assumed that the long-time phase function needs far less precision than the long-time magnitude function for preserving signal intelligibility. This exercise examines this assumption and shows that the relative roles of spectral magnitude and spectral phase are significantly different than expected.

This exercise selects a speech file and performs a long-time spectral analysis, where the long-term analysis computes a single FFT using the entire speech signal, and decomposes the resulting spectrum into magnitude and phase components. The exercise then re-creates the following three synthesized signals:

1. the speech signal obtained from retaining the original signal long-time phase, and substituting a unity magnitude contour in place of the computed long-time signal magnitude
2. the speech signal obtained from retaining the original signal long-time magnitude, and substituting a zero phase contour in place of the computed long-time signal phase
3. the speech signal computed from retaining the original signal long-time magnitude, and substituting a random phase contour in place of the computed long-time signal phase

By playing out the four signals (the original signal and the three re-synthesized signals), and plotting the original and re-synthesized signals on a series of plots, the user can compare and contrast the re-synthesized signals and think about the potential implications of the results on speech analysis/synthesis systems.

Spectral Magnitude/Phase – GUI Design

The GUI for this exercise consists of two panels, 4 graphics panels, 1 title box and 8 buttons. The functionality of the two panels is:

1. one panel for the graphics display,
2. one panel for selecting the speech file for analysis, and for running the program.

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

1. the original speech waveform on a normalized amplitude scale,
2. the signal obtained by re-synthesis of long-time phase with the long-time magnitude set to 1.0,
3. the signal obtained by re-synthesis of long-time magnitude with the long-time phase set to 0,
4. the signal obtained by re-synthesis of long-time magnitude with the long-time phase set to a random variable.

The title box displays the information about the selected file for analysis of long-time features. The functionality of the 4 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 run the exercise,
4. a pushbutton to close the GUI,
5. a pushbutton to play the original speech file,
6. a pushbutton to play the file with original phase and magnitude set to 1.0,
7. a pushbutton to play the file with original magnitude and phase set to 0,
8. a pushbutton to play the file with original magnitude and random phase.

Spectral Magnitude/Phase – Scripted Run

A scripted run of the program 'phase_magnitude_GUI25.m' is as follows:

1. run the program 'phase_magnitude_GUI25.m' from the directory 'matlab_gui\phase_magnitude_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 'we were away a year ago_suzanne.wav' for this example,
4. hit the 'Run Phase Magnitude' button to compute and display the results of long-time re-synthesis of the speech signal with variants on the long-time magnitude and phase; the original long-time speech signal is shown in the upper graphics panel; the re-synthesized signal with original long-time phase but with long-time magnitude set to 1 is shown in the second graphics panel; the re-synthesized signal with original long-time magnitude but with long-time phase set to 0 is shown in the third graphics panel; and the re-synthesized signal with original long-time magnitude but with long-time phase set to a Gaussian random value is shown in the fourth graphics panel; the user can listen to each of these files by clicking OK on the text message box that appears for each of these four signals,
5. hit the four buttons associated with playing the four versions of the speech file in any desired order,
6. experiment with a different choice of speech file,
7. 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_suzanne.wav'

is shown in Figure 1. The graphics panels show the original speech waveform (top graphics panel), the synthesized utterance using the original long-time phase but with the long time magnitude set to 1.0 in the second graphics panel; the synthesized utterance using the original long-time magnitude but with the phase set to 0 in the third graphics panel; and the synthesized utterance using the original long time magnitude with random phase, in the bottom graphics panel. A cursory glance at these four plots shows that the only case where the intelligibility is preserved is when using the original phase; all the other cases lead to degradations that make the speech unintelligible.

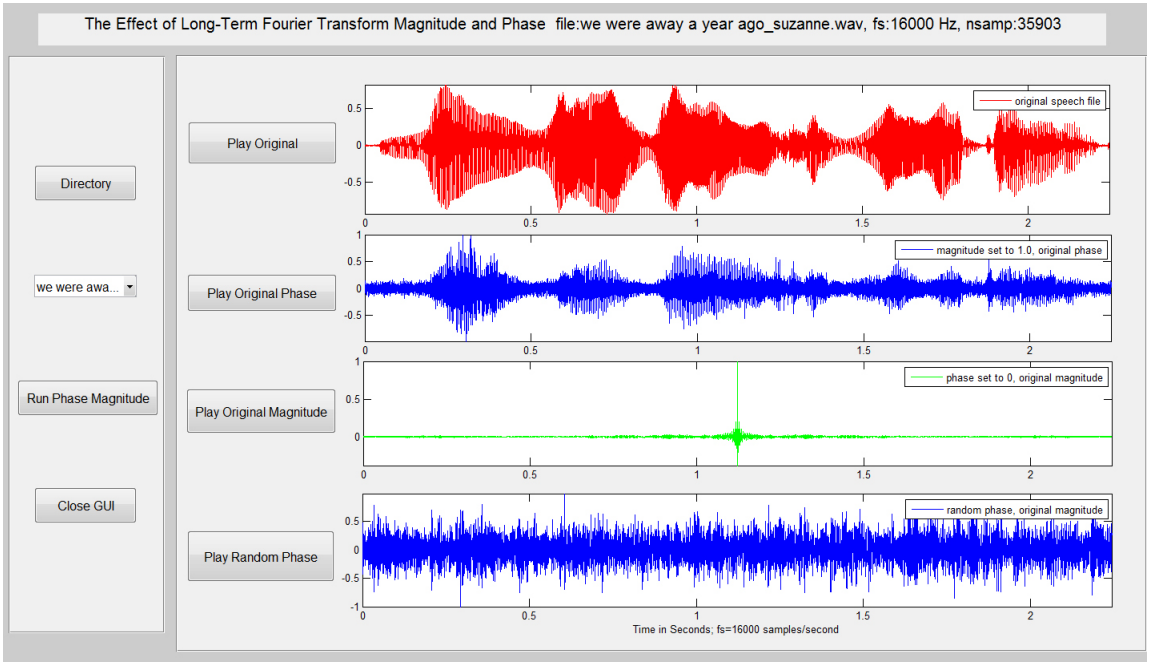

Figure 1: Long-time analysis of the speech file 'we were away a year ago_suzanne.wav'. Upper graphics panel shows original speech file waveform; second graphics panel shows re-synthesis with original phase and magnitude set to 1; third graphics panel shows re-synthesis with original magnitude and phase set to 0; fourth graphics panel shows re-synthesis with original magnitude and random phase

Spectral Magnitude/Phase – Issues for Experimentation

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

   - what does the synthesis with original long-time phase, and with long-time magnitude set to 1.0 sound like?
   - what does the synthesis with original long-time magnitude, and with long-time phase set to 0 sound like?
   - what does the synthesis with original long-time magnitude, and with random phase sound like?
   - can you explain why preserving long-time phase appears to be more important than preserving long-time magnitude?
   - why is the magnitude only version of the signal symmetric about the center of the graph?

2. run this exercise using the speech file 's5_edited.wav'

   - do the same effects occur with this new file; i.e., does it appear to be more important for speech intelligibility to preserve long-time phase than to preserve long-time magnitude?