Pattern and Speech Recognition WS2015-16
Exercise 2

Atanas Poibrenski(2554135), Marimuthu Kalimuthu(2557695), Furkat Kochkarov(2557017)

November 14, 2015

1 Preparation

1.1 Loading and Plotting in the waveform
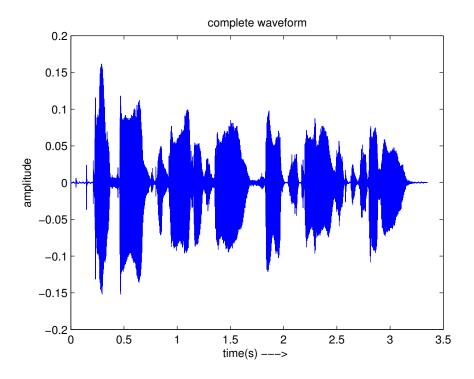
Ex.2 Complete waveform.

Figure 1: Complete waveform

Ex.3 Yes the same pattern is repeated multiple times in the signal. It roughly resembles the sawtooth waveform.
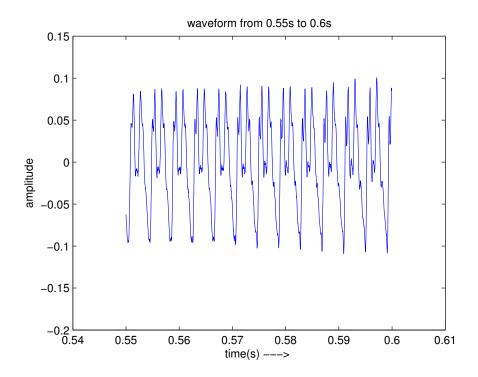
Figure 2: Waveform from 0.55s-0.6s

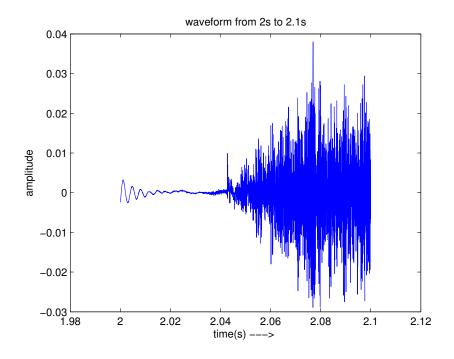**Ex.4** No pattern can be found in the signal.

Figure 3: Waveform from 2s-2.1s

2 Spectrum Analysis

2.1 A reference plot

Ex.5 A spectrogram shows how the frequency content of a signal changes over time. It's a 2-dimensional function of amplitude (brightness or color) vs frequency (vertical axis) vs time (horizontal axis)
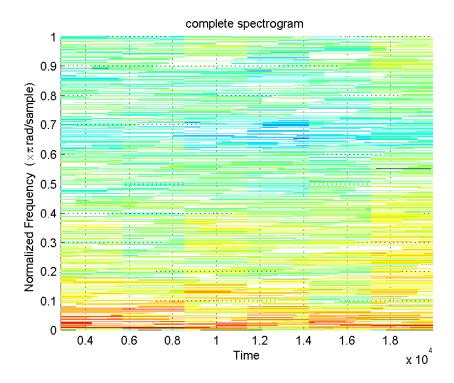

Figure 4: Spectrogram of the sample

Ex.6 It is a voiced segment.

Reason: When producing voiced speech, the air exhaling out of lungs through trachea will be interrupted periodically by the vibrating vocal folds. Because of this, the glottal wave is generated that excites the speech production system resulting in voiced speech. Thus, when we observe at a speech signal waveform, if it looks *nearly periodic in nature*, then it can be marked as voiced speech.
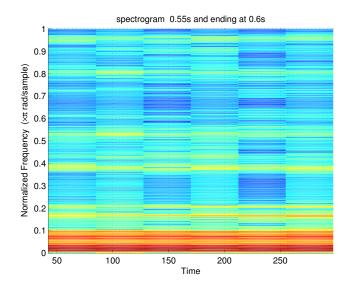
Figure 5: Spectrogram of the segment from 0.55s-0.6s

Ex.7 It is an unvoiced segment.

Reason: The waveform that we observe has no periodic pattern and is caused by some random noise.

During unvoiced speech duration, somewhere along the length of vocal tract, total or partial closure occurs resulting in obstruction of air flow completely or narrowly. This modification of airflow causes the vocal tract system to produce unvoiced speech.
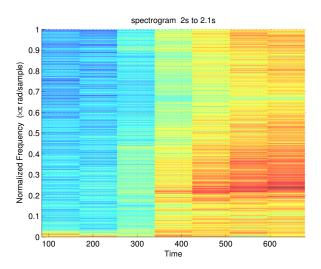

Figure 6: Spectrogram of the segment from 2s-2.1s

2.2 Compute spectrum

2.2.1 windowing

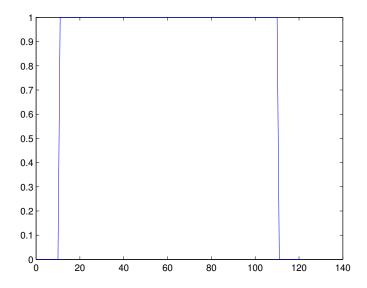**Ex.8**

- Rectangular window : See "rectWindow.m"

4

Figure 7: Rectangular window
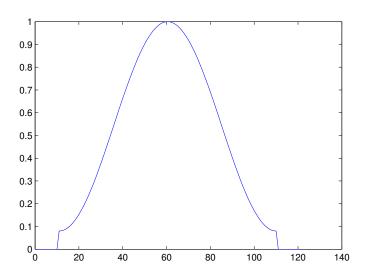
- Hamming window : See "hammingWindow.m"

Figure 8: Hamming window

Ex.9 It implies that we have an overlap of windows equivalent to the duration of frameshift. Here we will have an overlap of 5ms. We choose the hamming window because it produces much less spectral leakage.

Ex.10 See "applyHamming.m"

2.2.2 Generate a spectrum section

Ex.11

$$ss = \frac{frame\ start\ time}{frameshift} * \left\{ \frac{frameshift}{length\ of\ x(ms)} \right\} * (Number\ of\ samples\ in\ x)$$

$$= \frac{550}{5} * \frac{5}{3350} * 160800$$

$$= 26400$$

$$ee = ss + \left\{ \frac{winsize}{length\ of\ x(ms)} * (Number\ of\ samples\ in\ x) \right\}$$

$$= 26400 + \left\{ \frac{25}{3350} * 160800 \right\}$$

$$= 27600$$

So we have to use the index from 26400 to 27600.

Ex.12 Yes this result is expected because the signal is periodic and it is symmetric with respect to the center point of the x-axis. This represents the Nyquist frequency.
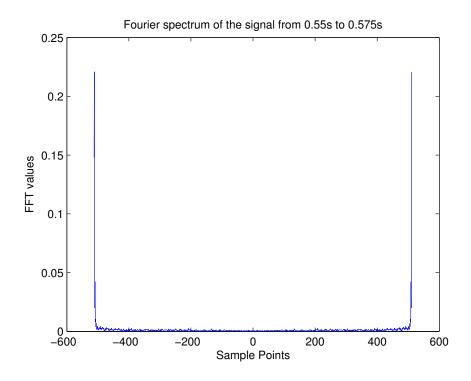

Figure 9: Fourier spectrum applied

Ex.13 It helps to visualize the fourier transform with the zero-frequency component in the middle of the spectrum.
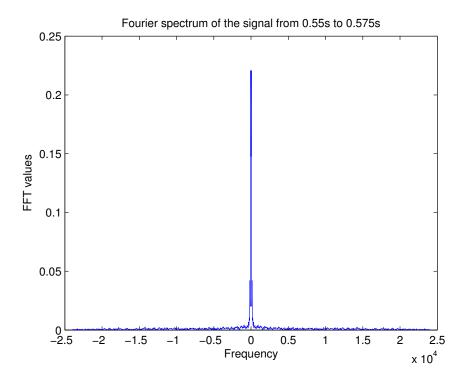
Figure 10: FFT shifted

Ex.14 Since the signal is symmetric(periodic) in the fourier domain, we can just remove one part of the symmetry and keep the other part. See Figure 9.

Ex.15 The *x-axis* should represent the frequency. In order to achieve this, we should normalize the indices and multiply them by the sampling rate. See Figure 10. Also see "ex02.m"(line:89)

2.2.3 Generate the power spectrum

Ex.16 See "powerSpectrum.m"
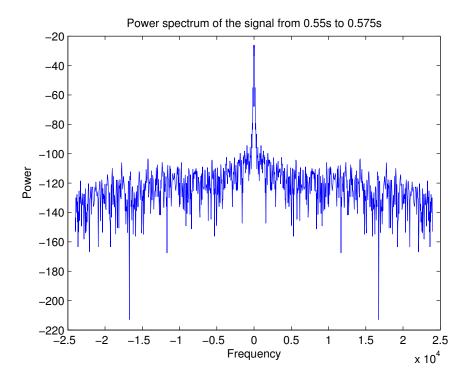
Ex.17 Power spectrum

Figure 11: Power spectrum

Ex.18 To plot a spectrogram we need *time interval* together with the frequencies and the corresponding amplitudes. The power spectrum can provide time interval, frequencies, and power per frequency.

Since power is proportional to the square of amplitude, we can compute the corresponding amplitudes for each frequency.

Bonus - From spectrum to mel-cepstrum
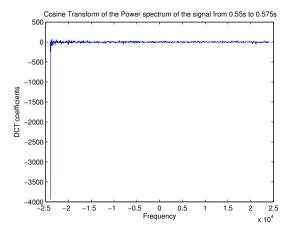
Ex.19 Applying discrete cosine transform.

Figure 12: DCT to 0.55s