Open lab –Speech/Audio signal Processing using MATLAB

Lab Sheet 3

Pre-processing of Speech

Aim

- Choosing sampling frequency for speech processing.
- Filter design using pole-zero plot
- Noise removal and Pre emphasis filtering

Theory

Speech signal processing on a digital machine needs sampling and storing of the analog version of the speech signal generated at the output of microphone. Sampling frequency is the parameter that controls the sampling process. The number of bits per sample is the parameter that controls the bit resolution. The intelligibility, amount of information and also the perceptual quality of speech depends on these two parameters.

Sampling theorem and sampling frequency

The sampling of analog signal is based on sampling theorem. The sampling theorem states that if fm is the maximum frequency component in the analog signal, then the information present in the signal can be represented by its sampled version provided the number samples taken per second is greater than or equal to twice the maximum frequency component. The number of samples/second is more commonly termed as sampling frequency fs. According to sampling theorem, fs should be greater than or equal to 2 fm.

The speech signal has frequency components in the audio frequency range (20 Hz to 20 kHz) of the electromagnetic spectrum. This is the reason for perceiving the information present in the speech signal by human ears. The fundamental question is up to what range of audio frequency, the speech signal has frequency components. We can analyze this experimentally by considering the whole audio range. The standard sampling frequency to sample the entire audio range is 44.1 kHz. This is because, 20 kHz is the maximum frequency component and allowing some guard band, the sampling frequency has been set at 44.1 kHz.

Lab Exercise

1. Let us consider the speech signal uploaded in AUMS and plot the spectra for the waveform
2. Segment the speech into selected different 30 msec segments and plot the spectra for each segment
3. Write about the average frequency band of the speech spectra.

Usually it is observed from different spectra that, there are no significant frequency components in the spectrum beyond about 4 kHz. This observation shows that 44.1 kHz sampling is too high value to capture the information present in the speech signal. Since information is up to about 4 kHz, 8 kHz is the minimum sampling frequency.

Pre-processing of Speech

1. Noise and DC removal

The speech samples while recording will be affected by noise. Also the spectra in and around zero frequencies are irrelevant viz. DC offset while recording. Hence a band-pass filtering has to be performed before we actually process the speech samples.

A sample butterworth bandpass filter specification is given below, with a lower cut-off of frequency of 700 Hz and upper cut off frequency 8KHz, and order 7 .

n = 7
Nyquist=fs/2
beginFreq = 700 / Nyquist ; endFreq = 8000 / Nyquist

Lab Exercise

4. Pass the signal specified in que.no1 through the filter specified above. Play the filtered signal.
5. Plot the spectra of the segment of speech you have taken and check whether the spectrum other than the specified band pass range is filtered?

2. Pre- emphasis filtering

Pre-emphasis is a very simple signal processing method which increases the amplitude of high frequency bands and decrease the amplitudes of lower bands. Pre-emphasis is a way of compensating for the rapid decaying spectrum of speech, which is due to the peculiar shape of the vocal tract. This is done using a high pass FIR filter given by the difference equation:

$$y(n) = x(n) - a\,x(n-1)$$

a is usually between 0.9 and 1.

Lab Exercise

6. For the FIR filter given above, plot the frequency response using the command `freqz` and verify whether it is a high pass filter.
7. Go for the pole-zero plot using the command `zplane`.
8. Comment on the location of poles and zeros and from this write about the approximate frequency response

9. Apply this filtering to the noise removed signal obtained and playback the signal and observe
10. Check the spectra of the signal you have finally obtained.

3. Frame blocking

Speech is produced from a time varying vocal tract system with time varying excitation. As a result the speech signal is non-stationary in nature. Most of the signal processing tools studied in signals and systems and signal processing assume time invariant system and time invariant excitation, i.e. stationary signal. Hence these tools are not directly applicable for speech processing. This is because, use of such tools directly on speech violates their underlying assumption.

Studies have shown that the vocal tract preserves its stationary nature for about 10-20 m sec while speaking, due to the inertia of facial muscles. Hence the input speech signal is segmented into frames of 20~30 ms with optional overlap of 1/3~1/2 of the frame size.

4. Windowing

Each frame has to be multiplied with a hamming window in order to keep the continuity of the first and the last points in the frame. If the signal in a frame is denoted by s(n), n = 0,…N-1, then the signal after Hamming windowing is s(n)*w(n), where w(n) is the Hamming window defined by:

$$w(n) = (1 - a) - a\cos\left(\frac{2pn}{N-1}\right), \quad 0 \leqq n \leqq N-1; a = 0.45$$

Assignment

1. Perform Que. No. 6 ,7 and 8 to the systems described below:

   (a) $y(n) = x(n) + x(n-1)$
   (b) $y(n) = [x(n) + x(n-1) + x(n-2)]/3$

2. Create a matlab script to segment the speech automatically into selected different 30 msec segments
3. In the segmented speech, apply Hamming window and plot each frame.