Digital Signal Processing Laboratory

Experiment 6

By: Hardik Tibrewal (18EC10020)

Speech Recognition

Objective:

Speech recognition through primarily temporal cues.

Theoretical Background: (From the paper by Robert V. Shannon et al.)

The recognition of speech has been primarily thought of as frequency specific, since spectral energy peaks in speech reflect the resonant properties of the vocal tract and thus provide acoustic information on the production of the speech sound. However, using spectral cues identify phonemes were met with limited success. This is why temporal cues were also used, which contain a lot of information due to their highly complex characteristics. Even total removal of spectral cues from speech resulted in stimuli with a high amount of information. By preserving amplitude and temporal cues while varying the amount of spectral information, we can parametrically assess the role of spectral detail in speech recognition independently of temporal cues.

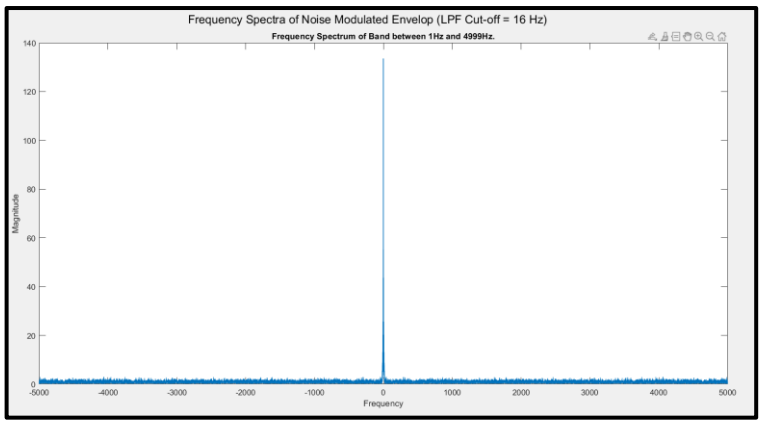
Pseudocode:

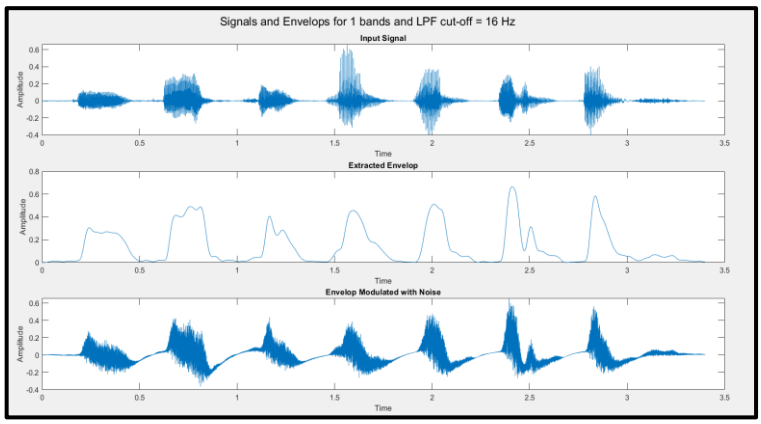
The audio file was first upsampled from 8 kHz sampling frequency to a 10 kHz sampling rate. This was done since the paper uses a 10 kHz sampling rate and the frequency band cut-offs were decided by this. The same processing can be done on the signal sampled at 8 kHz as well. Then the signal was divided into several frequency bands (1-4 bands) using bandpass filters. The cut-off frequencies of these bands were taken from the paper itself, and the lowest and highest frequency is 1 Hz and 4999 Hz. This is so that the digital filter in each loop can be created using the bandpass filter function, rather than handling separate cases for a low-pass and high-pass filter. The amplitude envelope was extracted from each band through half-wave rectification and low-pass filtering. The cut-off frequency of the LPF

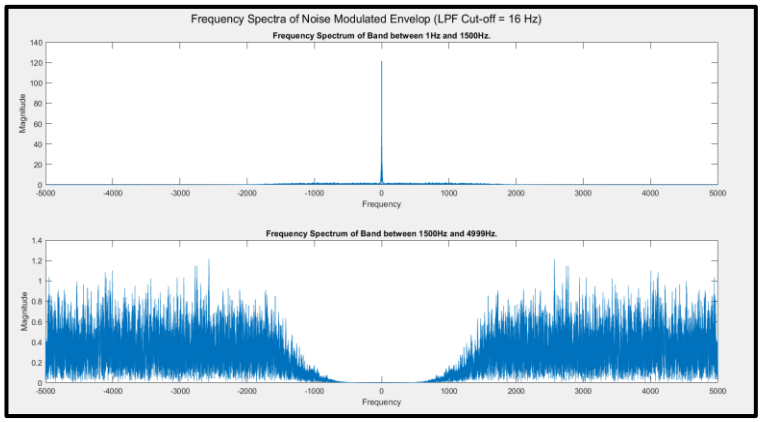
used for envelope extraction was also varied between the values mentioned in the paper (16 Hz, 50 Hz, 160 Hz, and 500 Hz). The cut-off of the LPF was varied to evaluate the effect of reducing the bandwidth of temporal envelope information. The envelope signal was then used to modulate white noise, which was then spectrally limited by the same bandpass filters used to split the original speech signal into frequency bands. Finally, all bands were summed, passed through a low-pass filter of cut-off 4 kHz, and then the output was amplified and written into an audio file. The envelope extracted from the signal was also amplified and written in a file.

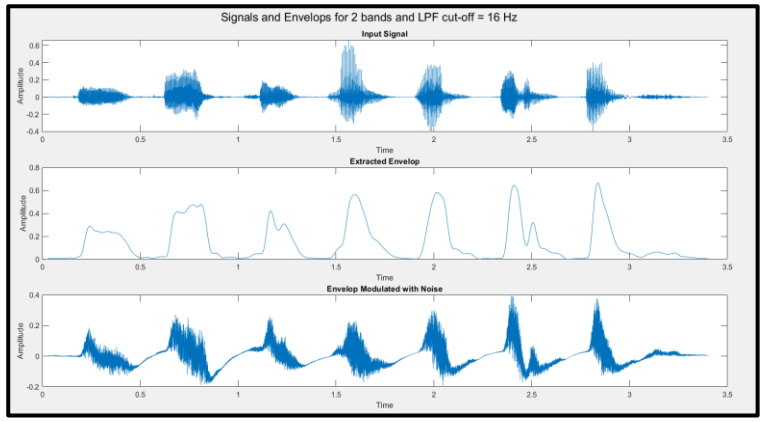
The frequency domain and time domain plots of the signals were plotted for each of the 16 cases. The output audio files were compared to the original input and the ease of recognition of the original message from the processed output was noted.

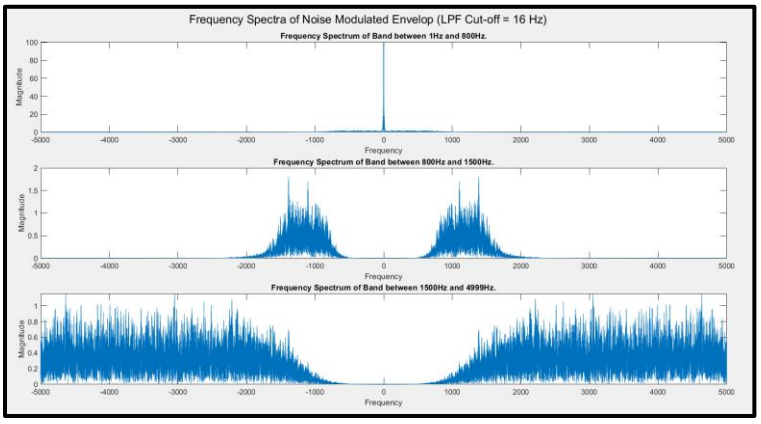
Simulation Results:

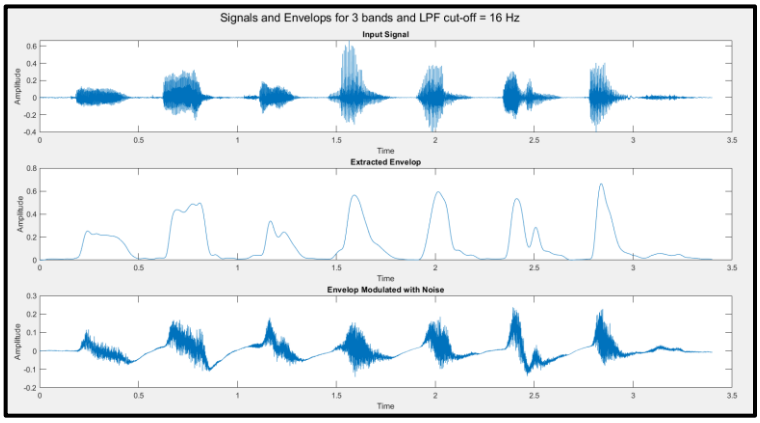


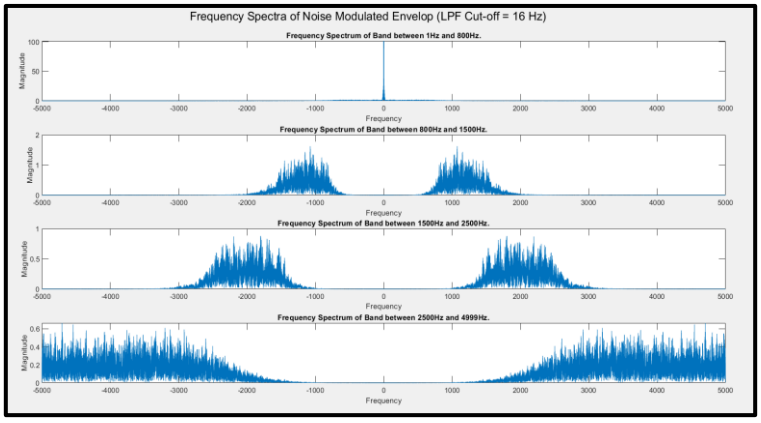


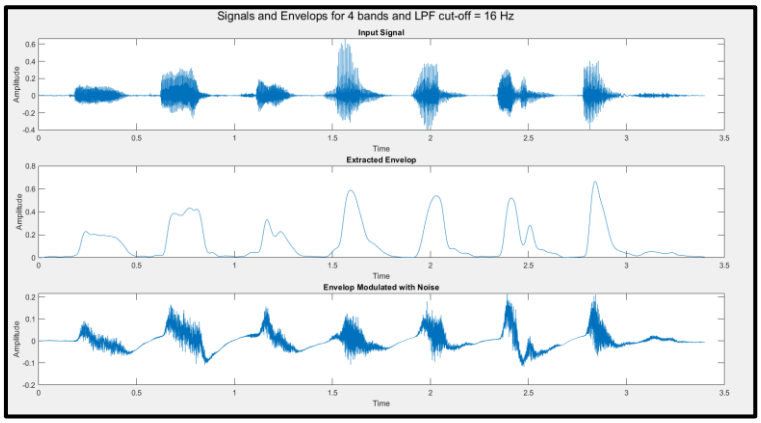


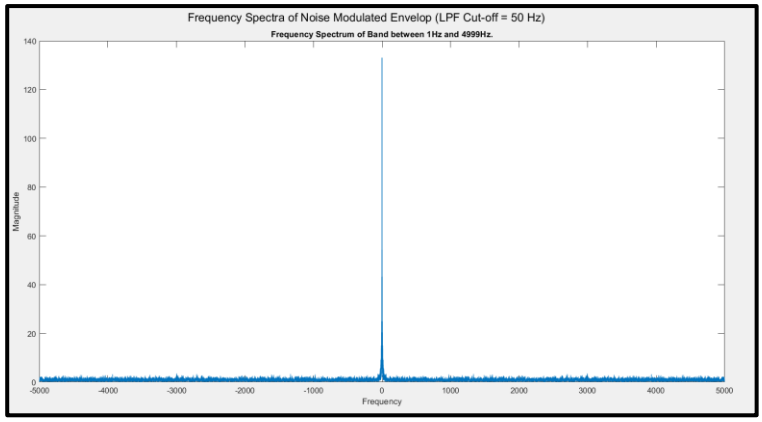


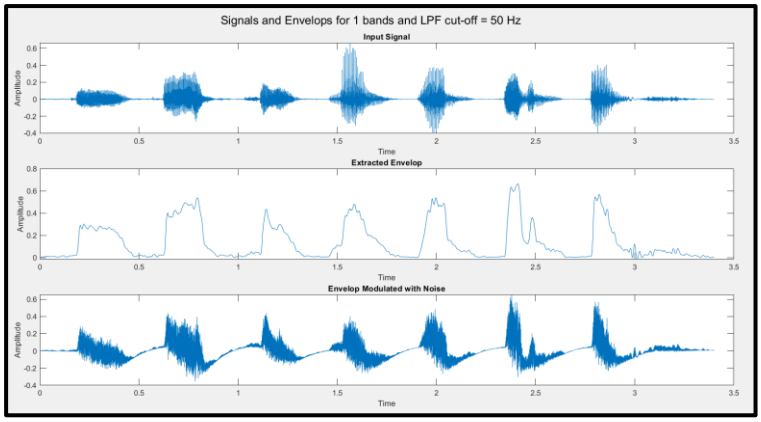


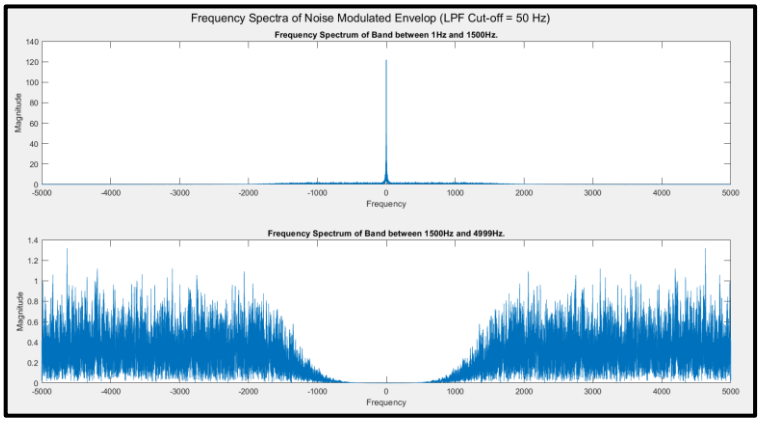


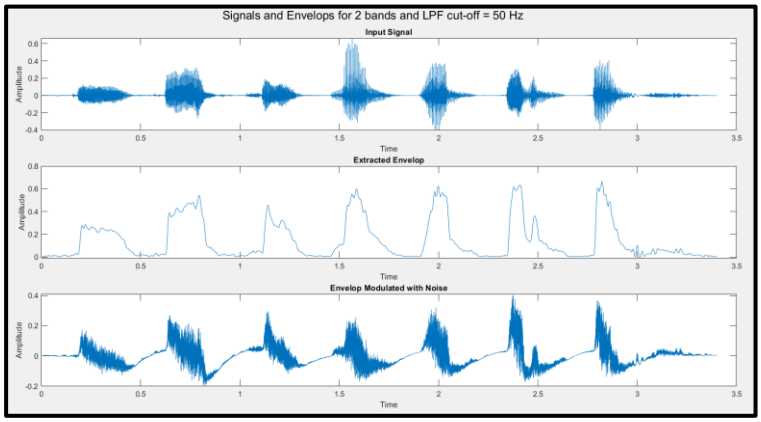


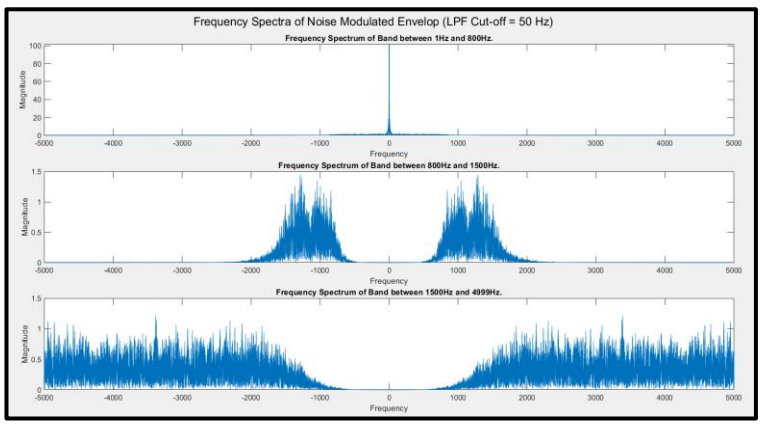


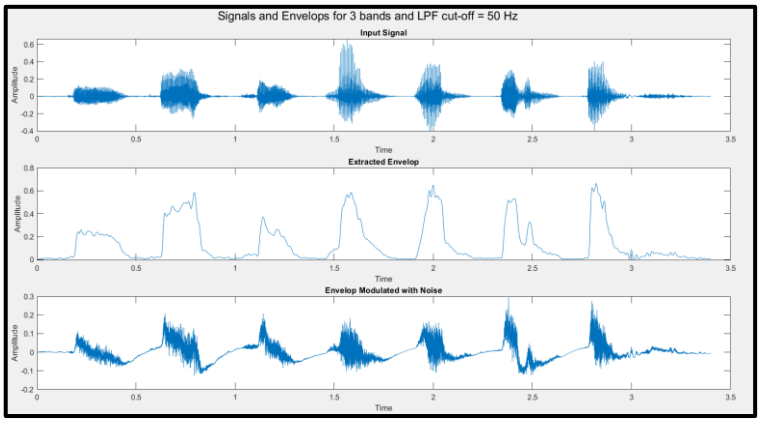

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 50 Hz)
Frequency Spectrum of Band between 1Hz and 4999Hz.
Magnitude
Frequency

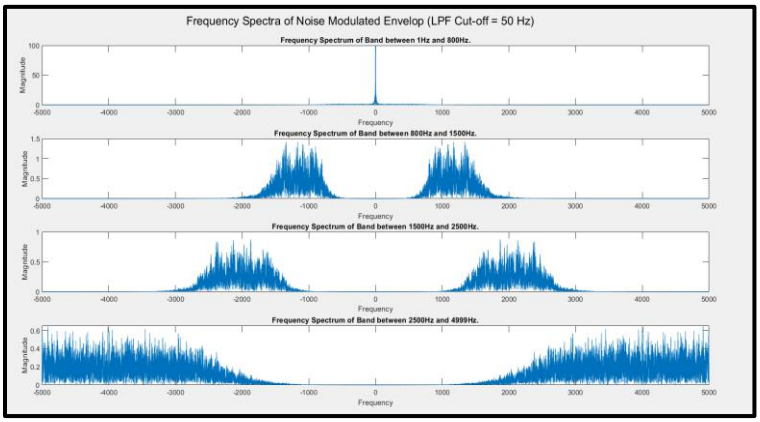

Signals and Envelops for 1 bands and LPF cut-off = 50 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

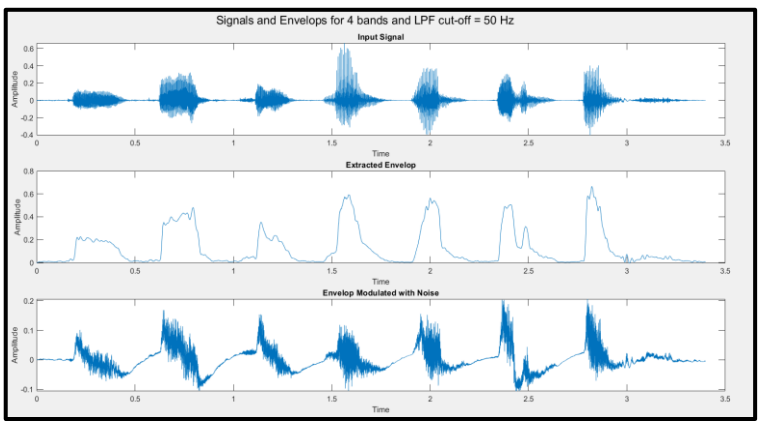

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 50 Hz)
Frequency Spectrum of Band between 1Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 4999Hz.
Magnitude
Frequency

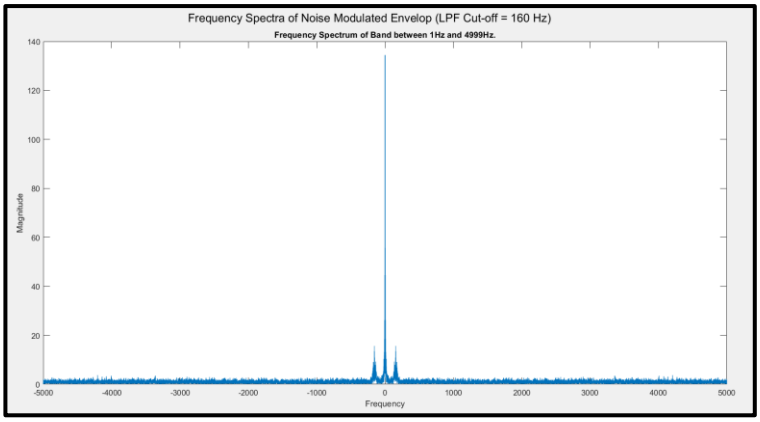

Signals and Envelops for 2 bands and LPF cut-off = 50 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

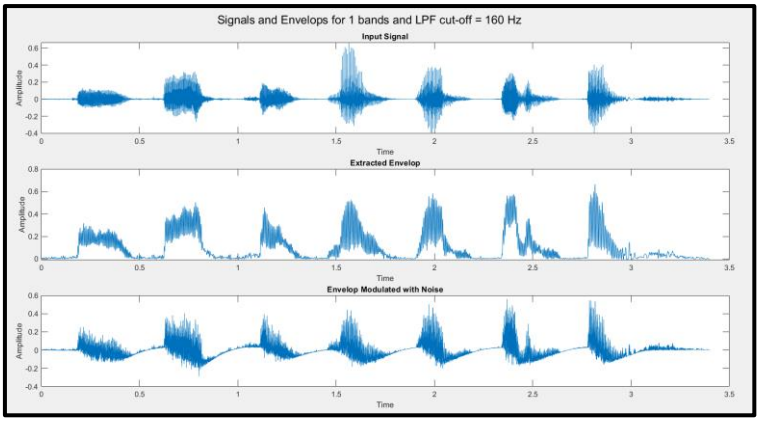

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 50 Hz)
Frequency Spectrum of Band between 1Hz and 800Hz.
Frequency Spectrum of Band between 800Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 4999Hz.
Magnitude
Frequency

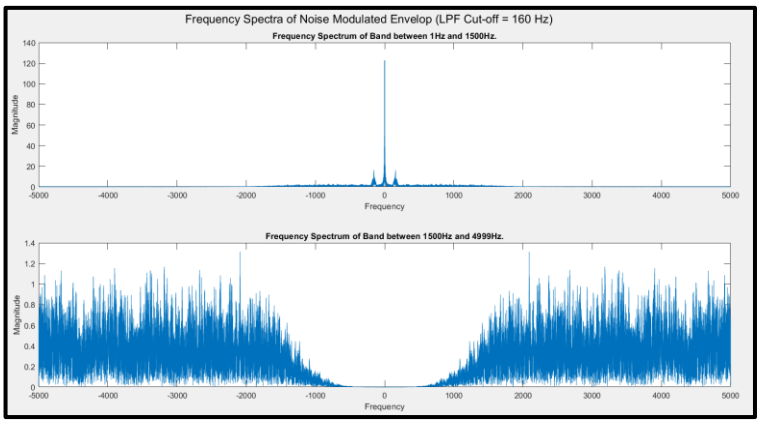

Signals and Envelops for 3 bands and LPF cut-off = 50 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

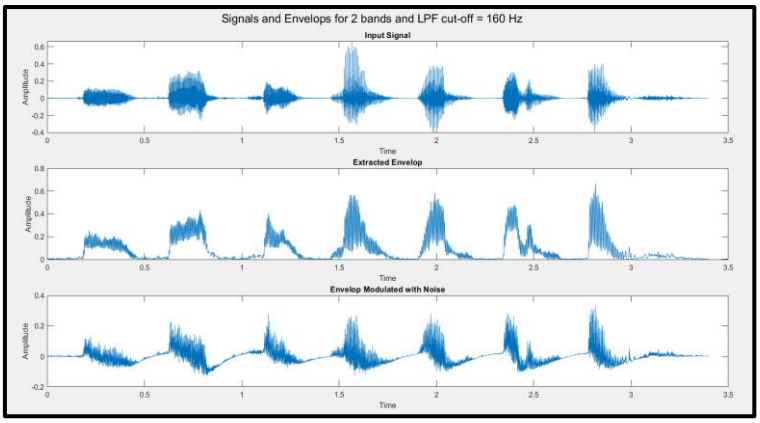

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 50 Hz)
Frequency Spectrum of Band between 1Hz and 800Hz.
Frequency Spectrum of Band between 800Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 2500Hz.
Frequency Spectrum of Band between 2500Hz and 4999Hz.
Magnitude
Frequency

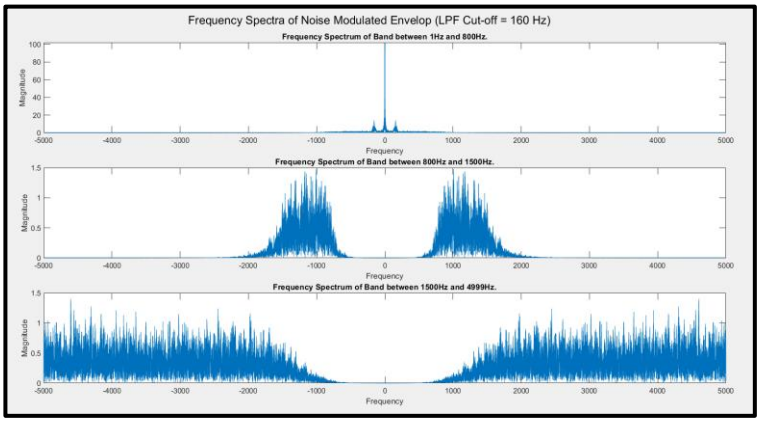

Signals and Envelops for 4 bands and LPF cut-off = 50 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

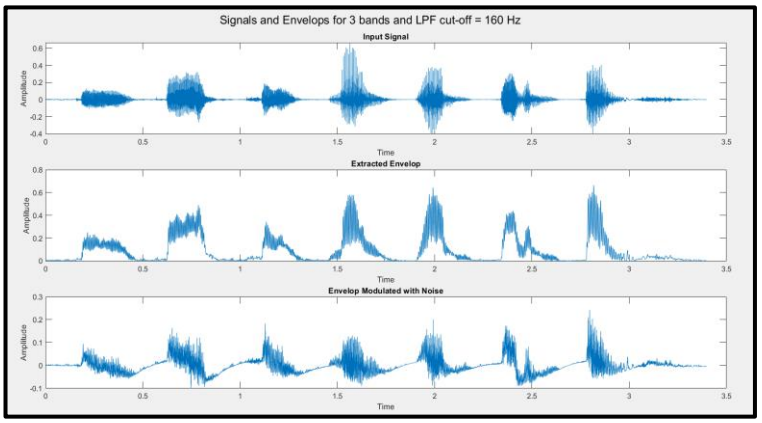

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 160 Hz)
Frequency Spectrum of Band between 1Hz and 4999Hz.
Magnitude
Frequency

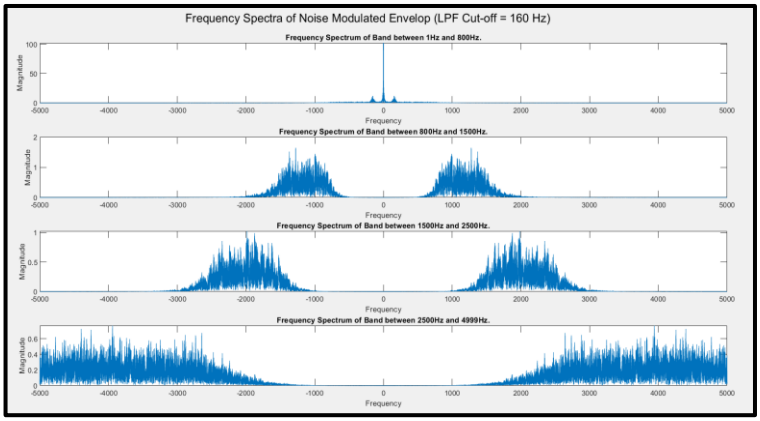

Signals and Envelops for 1 bands and LPF cut-off = 160 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

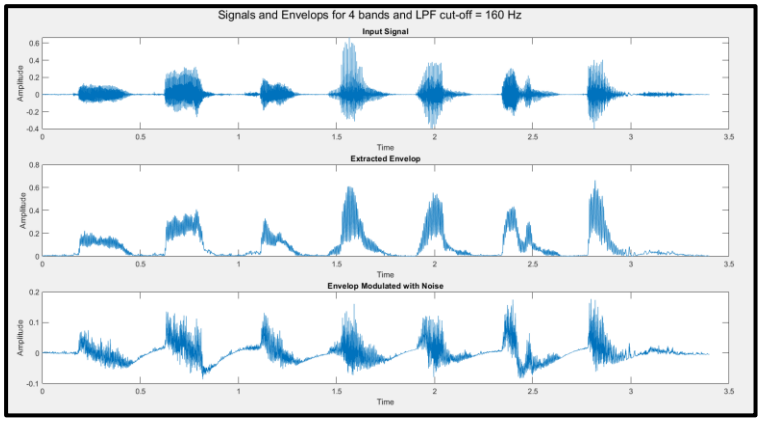

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 160 Hz)
Frequency Spectrum of Band between 1Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 4999Hz.
Magnitude
Frequency

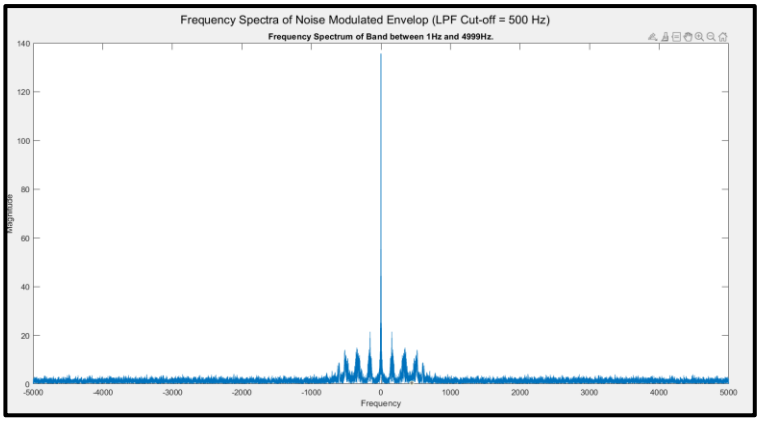

Signals and Envelops for 2 bands and LPF cut-off = 160 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

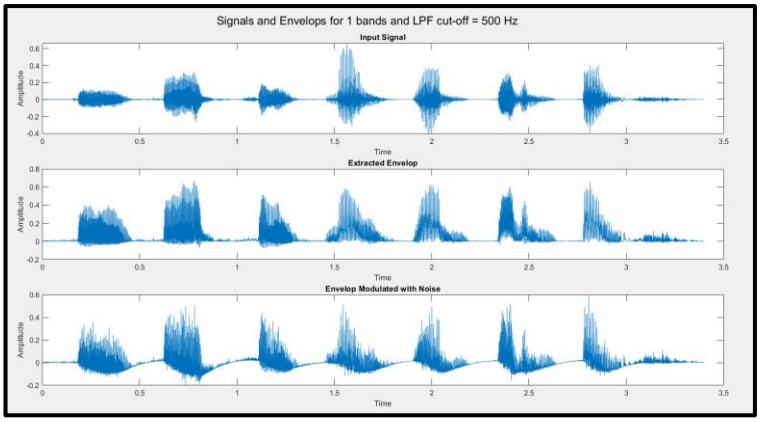

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 160 Hz)
Frequency Spectrum of Band between 1Hz and 800Hz.
Frequency Spectrum of Band between 800Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 4999Hz.
Magnitude
Frequency

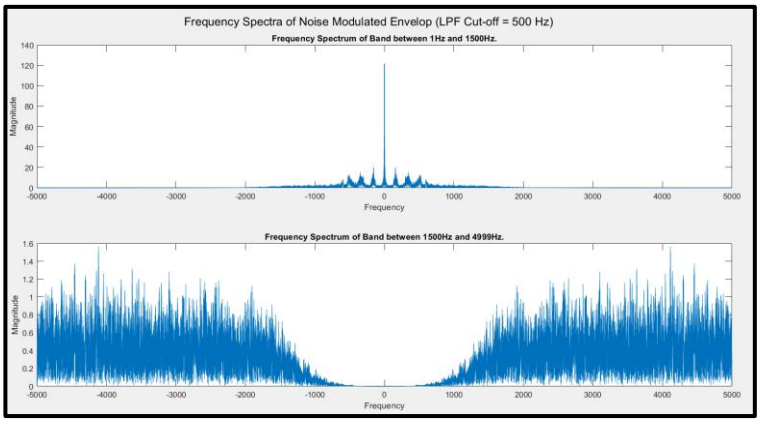

Signals and Envelops for 3 bands and LPF cut-off = 160 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

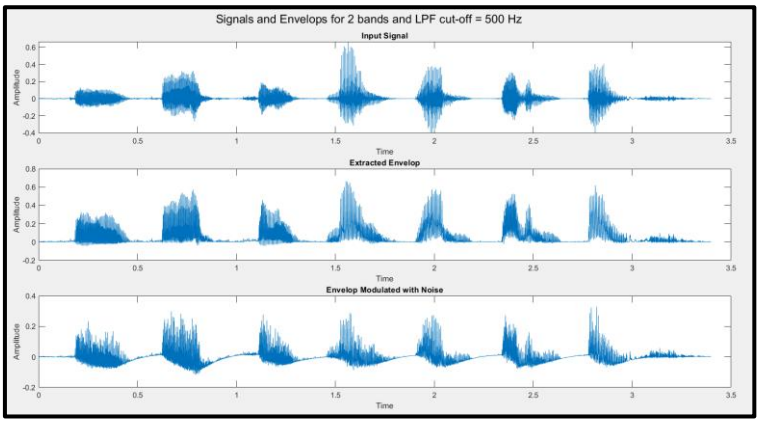

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 160 Hz)
Frequency Spectrum of Band between 1Hz and 800Hz.
Frequency Spectrum of Band between 800Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 2500Hz.
Frequency Spectrum of Band between 2500Hz and 4999Hz.
Magnitude
Frequency

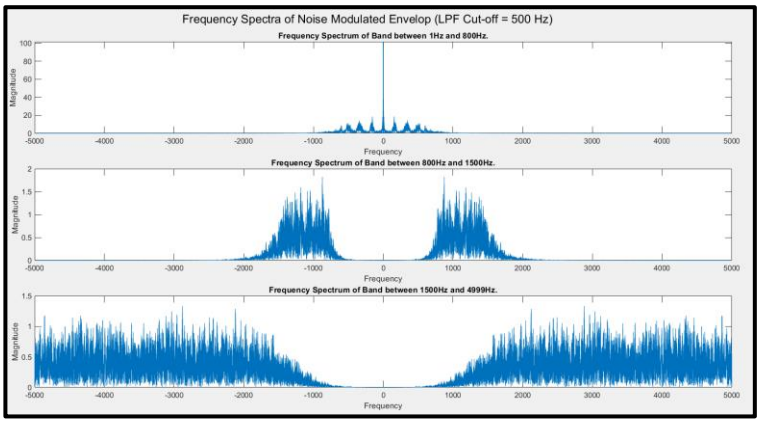

Signals and Envelops for 4 bands and LPF cut-off = 160 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

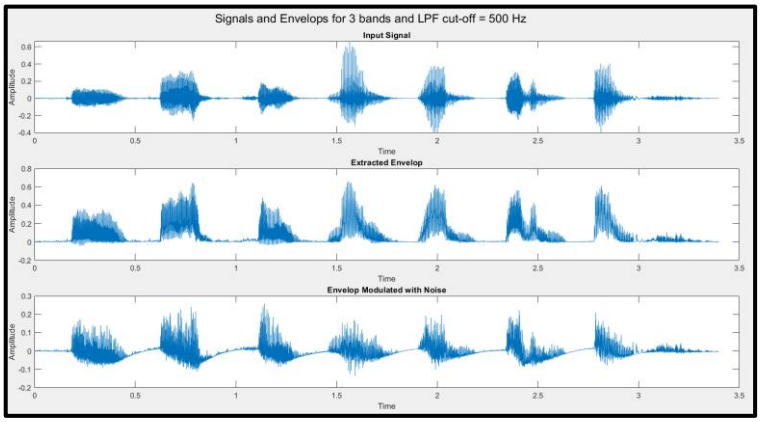

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 500 Hz)
Frequency Spectrum of Band between 1Hz and 4999Hz.
Magnitude
Frequency

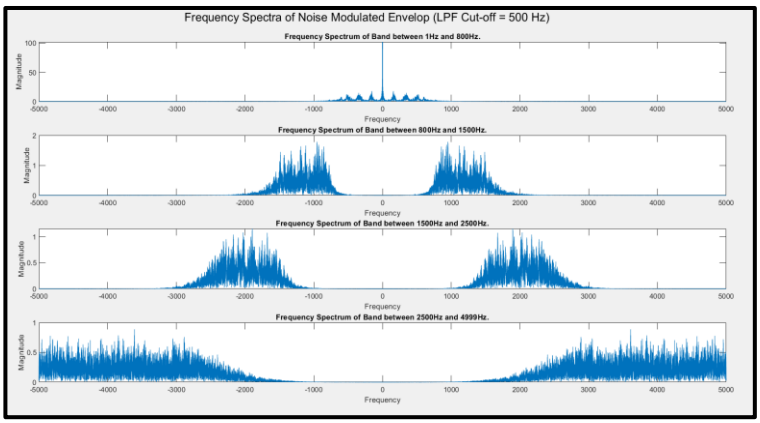

Signals and Envelops for 1 bands and LPF cut-off = 500 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

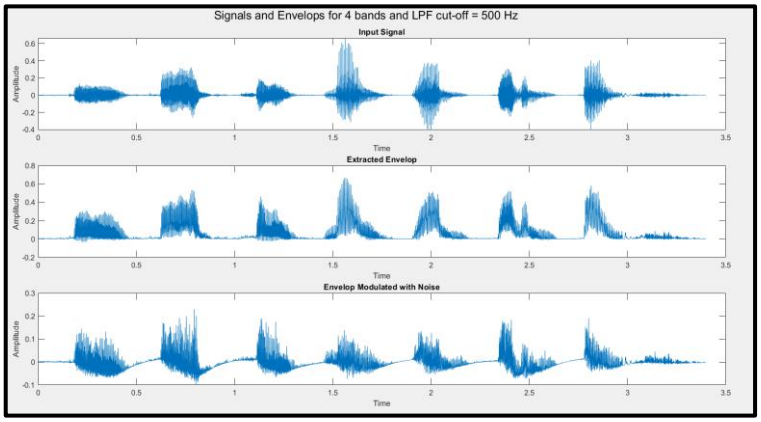

Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 500 Hz)
Frequency Spectrum of Band between 1Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 4999Hz.
Magnitude
Frequency


Signals and Envelops for 2 bands and LPF cut-off = 500 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time


Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 500 Hz)
Frequency Spectrum of Band between 1Hz and 800Hz.
Frequency Spectrum of Band between 800Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 4999Hz.
Magnitude
Frequency


Signals and Envelops for 3 bands and LPF cut-off = 500 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time


Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = 500 Hz)
Frequency Spectrum of Band between 1Hz and 800Hz.
Frequency Spectrum of Band between 800Hz and 1500Hz.
Frequency Spectrum of Band between 1500Hz and 2500Hz.
Frequency Spectrum of Band between 2500Hz and 4999Hz.
Magnitude
Frequency


Signals and Envelops for 4 bands and LPF cut-off = 500 Hz
Input Signal
Extracted Envelop
Envelop Modulated with Noise
Amplitude
Time

Audio Files: Audio Files

Results:

The envelope of the input audio was extracted and used to modulate the noise, which resulted in a recreation of the original speech signal, proving that temporal cues play a huge part in speech recognition. The clarity and comprehensibility of the modulated noise depended on the number of frequency bands and the cut-off frequency of the LPF used for envelope extraction, with a greater number of bands and higher cut-off frequency giving better results.

Discussion:

The input audio was processed to extract the speech envelope for 16 cases, by varying the number of frequency bands and the cut-off frequency of the envelope extraction filter. It was observed that an increase in both these values led to an audio output of better quality.

The reason for this is that more information is captured from the input signal when the processing increases, and it is a compromise between a better output and processing load. The envelope loses most of the spectral information as it limits the signal to a very low frequency range. However, this results in a very clear audio for the LPF cut-off of 500 Hz, and a faint but intelligible audio for a cut-off of 160 Hz. 16 Hz and 50 Hz cut-off frequencies resulted in no audible output, which is why the low frequency signals are modulated with band-limited higher frequency noise.

While the claim of the paper is verified that temporal cues contain a lot of information, more amount of spectral information is still beneficial, since a better audio quality of obtained when the number of bands was increased.

From the plots of the envelopes, we can see the effect of the LPF cut-off on the shape of the obtained envelope. Increasing the cut-off while keeping the number of bands constant resulted in a better output audio quality, as expected, since more information about the original spectral characteristics will be retained. However, the lesser, but significant, amount of information present in the temporal cues is still enough to understand the output in most cases.

Appendix

Code:

```
clc
clear all
close all

order = 4; % order of all filters
[y,F] = audioread("B1_A1.wav");
audiowrite("Original.wav",y, 10000); %same as upsampling
[y,Fs] = audioread("Original.wav"); %reading the original speech signal
norm = Fs/2;

for N = [1,2,3,4] %Number of frequency bands
   for Fc = [16, 50, 160, 500] %Cutoff frequency (Hz) of envelope extraction filter
      [B_l, A_l] = butter(order, Fc/norm);
      noise = rand(size(y));
      output = zeros(size(y));
      envelope = zeros(size(y));

      %Band cut-off frequencies as decided by the paper
      bands = zeros(4,5);
      bands(1,:) = [1, norm-1, 0, 0, 0];
      bands(2,:) = [1, 1500, norm-1, 0, 0];
      bands(3,:) = [1, 800, 1500, norm-1, 0];
      bands(4,:) = [1, 800, 1500, 2500, norm-1];

      figure();
      sgtitle("Frequency Spectra of Noise Modulated Envelop (LPF Cut-off = "+Fc+" Hz)")
      for ii = 1:N
         [B, A] = butter(order, [bands(N,ii)/norm, bands(N,ii+1)/norm]);
         Y = filter(B,A,y); %Creating frequency bands
         Y_e = Y.*(Y>=0); %Half-wave rectification
         Y_el = filter(B_l, A_l, Y_e); %Envelope extraction by LPF
         noise_mod = noise.*Y_el; %Noise-modulation
         noise_mod = filter(B,A,noise_mod); %Spectral limitation

         subplot(N,1,ii);
         NUM = length(Y_el);
         f_range = -norm:2*norm/NUM:norm-1/NUM;
         plot(f_range, abs(fftshift(abs(fft(noise_mod))))); %Frequency spectrum of band
         xlabel("Frequency");ylabel("Magnitude");
         title("Frequency Spectrum of Band between "+bands(N,ii)+"Hz and "+bands(N,ii+1)+"Hz.");

         %Summing modulated noise and envelope for each band
         output = output + noise_mod;
         envelope = envelope+Y_el;
      end
      %Passing output through LPF of cutoff 4kHz and amplifying
      [B_f, A_f] = butter(order, 4000/norm);
      output = filter(B_f, A_f, output);
      output = output.*(max(y)/max(envelope));
      %Passing envelope through LPF of cutoff 4kHz and amplifying
      envelope = filter(B_f, A_f, envelope);
      envelope = envelope.*(max(y)/max(envelope));
      t = 0:1/Fs:(length(y)-1)/Fs;

      %Plotting all signals in time domain
```

```
        figure();
        sgtitle("Signals and Envelopes for "+N+" bands and LPF cut-off = "+Fc+" Hz");
        subplot(311)
        plot(t, y);
        xlabel("Time");ylabel("Amplitude");
        title("Input Signal");

        subplot(312)
        plot(t, envelope);
        xlabel("Time");ylabel("Amplitude");
        title("Extracted Envelop");

        subplot(313);
        plot(t,output);
        xlabel("Time");ylabel("Amplitude");
        title('Envelop Modulated with Noise');

        %Writing output and envelope as audio files
        out_file = "./Audio_outputs/answer_"+N+"_freq_"+Fc+".wav";
        audiowrite(out_file,output,Fs);

        out_file = "./Audio_outputs/envelope_"+N+"_freq_"+Fc+".wav";
        audiowrite(out_file,envelope,Fs);
    end
end
```