Digital Signal Processing Lab

Experiment 6 – Speech Recognition with Primarily Temporal Cues

Utkarsh Patel (18EC30048)

---

1 Objective

To understand the relative importance of the low-frequency temporal structure of speech and frequency content of speech in speech perception.

2 Theory

- This experiment is based on the work of Shannon et al. (1995), in which nearly perfect speech recognition was observed under conditions of significantly reduced spectral information.
- In the process, temporal envelopes of speech were extracted from broad frequency bands and were used to modulate noises of the same bandwidths.
- This manipulation preserved temporal envelope cues in each band but restricted the listener to severely degraded information on the distribution of spectral energy.
- It was observed by the authors that the identification of consonants, vowels, and words in simple sentences improved markedly as the number of bands increased.
- Thus, the authors concluded that the presentation of a dynamic temporal pattern in only a few broad spectral regions is sufficient for the recognition of speech.

3 Procedure

1. The original audio signal is digitalized at 8 KHz sampling rate.
2. The signal was then split into frequency bands depending upon the total number of filters used (which can be 1, 2, 3 or 4).
3. The filters were designed with following description:
    a. **Case 1:** When only single filter is used
        i. Filter 1: Low-pass filter with cut-off frequency 4 KHz
    b. **Case 2:** When two filters are used
        i. Filter 1: Low-pass filter with cut-off frequency 1.5 KHz
        ii. Filter 2: Band-pass filter with passband 1.5 KHz – 4 KHz
    c. **Case 3:** When three filters are used
        i. Filter 1: Low-pass filter with cut-off frequency 800 Hz
        ii. Filter 2: Band-pass filter with passband 0.8 KHz – 1.5 KHz
        iii. Filter 3: Band-pass filter with passband 1.5 KHz – 4 KHz
    d. **Case 4:** When four filters are used
        i. Filter 1: Low-pass filter with cut-off frequency 800 Hz
        ii. Filter 2: Band-pass filter with passband 0.8 KHz – 1.5 KHz
        iii. Filter 3: Band-pass filter with passband 1.5 KHz – 2.5 KHz
        iv. Filter 4: Band-pass filter with passband 2.5 KHz – 4 KHz
4. After splitting signal into various bands, amplitude envelope is extracted for each band using Hilbert transform
5. The envelopes were then passed through low-pass filter with cut-off frequencies of 16 Hz, 50 Hz, 160 Hz and 500 Hz to evaluate the effect of reducing the bandwidth of temporal envelope extraction.
6. The envelope signal was then used to modulate white noise. This is done for each frequency band.
7. The modulated white noise associated with the frequency bands were summed over to produce output signal, which was then low-passed with cut-off frequency 4 KHz.
8. This final output signal is analyzed and compared with the original signal.

4 Results

The original audio signal used for the simulation is sampled at 8 KHz, with speaker uttering numbers' two four three nine one seven five'.

Original file: https://drive.google.com/file/d/1RSmuGJESWnk-y-Go-_vV-NWaSnrN2b_R/view?usp=sharing

4.1 Using one frequency band only

In this case, the signal is low-passed at 4 KHz, the amplitude envelope is extracted and used for modulating white noise. The output file can be accessed from:

Envelope cut-off frequency (in Hz)	*Link to the audio file*
16	https://drive.google.com/file/d/1d67ZIX8ZubH2CsExVCw0toW27Mxj1GBY/view?usp=sharing
50	https://drive.google.com/file/d/18lGuLHXcV5ytL7Mnf-FMXNXATTC0di6q/view?usp=sharing
160	https://drive.google.com/file/d/1LZ6cONGH2YHjYkuldFRMJfQOmQZSgwDw/view?usp=sharing
500	https://drive.google.com/file/d/174hM61hzxaAgTlkKoGWKUUgjv20vcXlG/view?usp=sharing

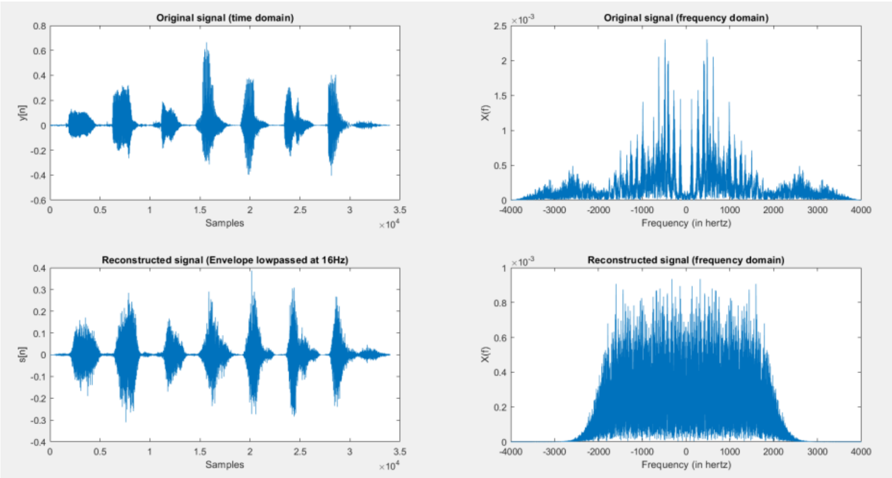

Fig. 1. ***Original V Reconstructed***: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 16 Hz

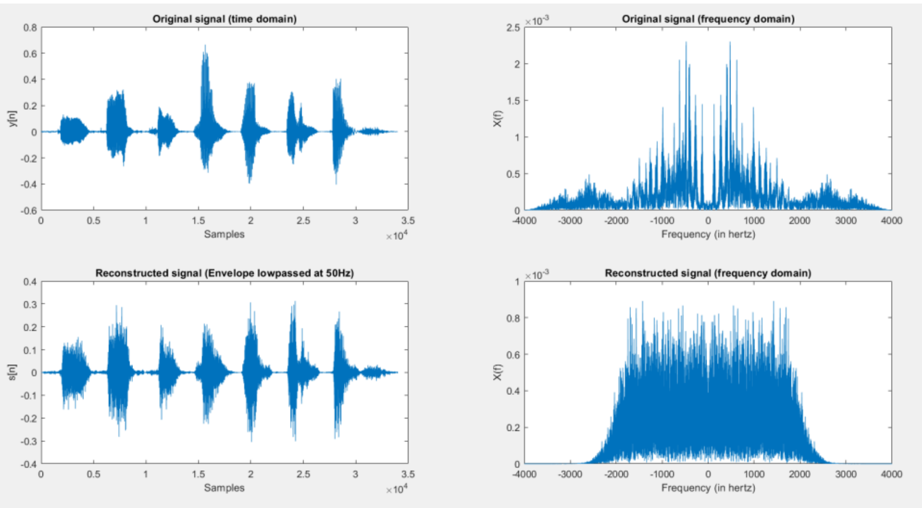

Fig. 2. ***Original V Reconstructed***: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 50 Hz

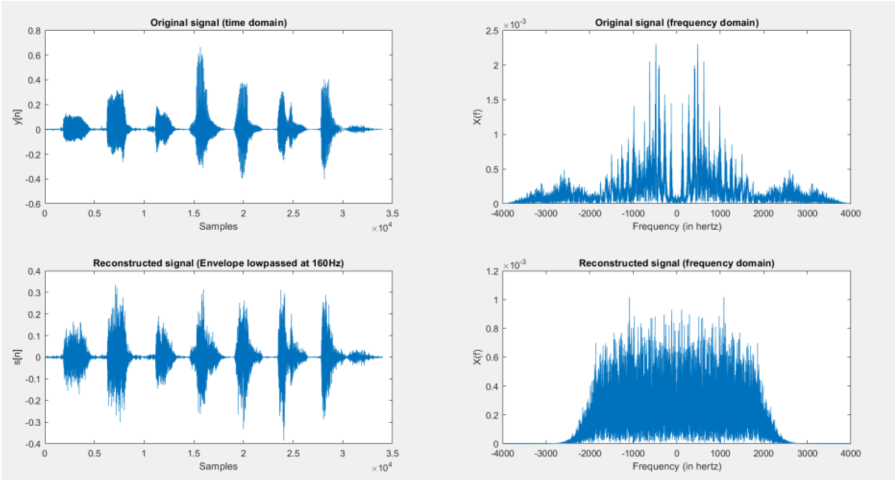

Fig. 3. ***Original V Reconstructed***: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 160 Hz

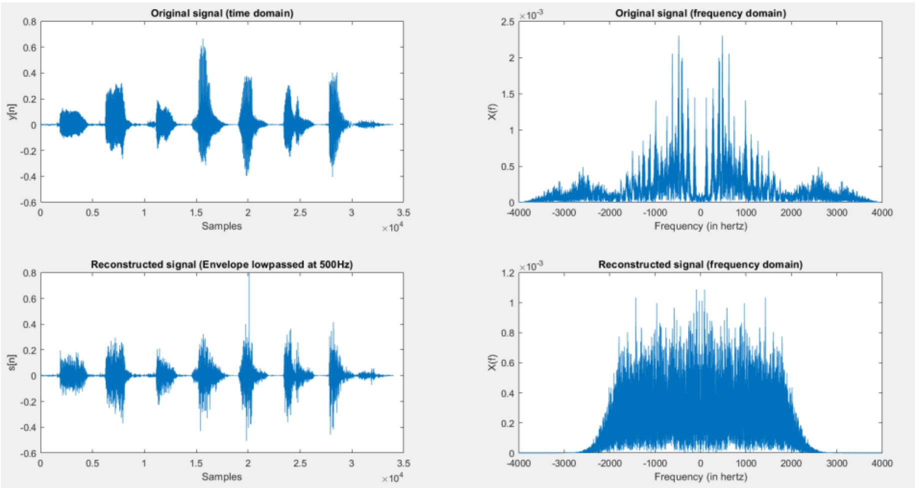

Fig. 4. ***Original V Reconstructed***: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 500 Hz

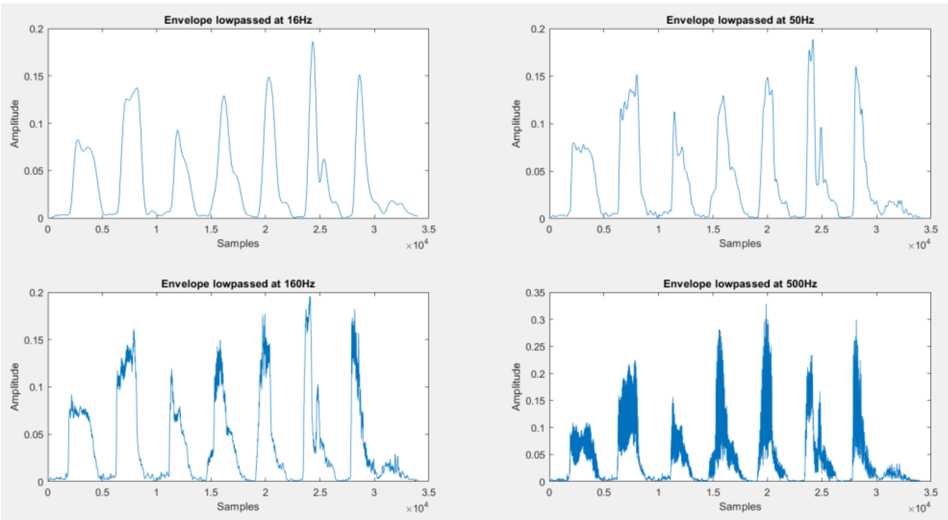

Fig. 5. ***Amplitude envelope***: For this plot, the amplitude envelope low-passed at 16 Hz, 50 Hz, 160 Hz and 500 Hz respectively when only one frequency band is used are shown

4.2 Using two frequency bands

In this case, the signal is using two filters:

1. Filter 1: Low-pass filter with cut-off frequency 1.5 KHz
2. Filter 2: Band-pass filter with passband 1.5 KHz – 4 KHz

The amplitude envelope is extracted and used for modulating white noise. The output file can be accessed from:

Envelope cut-off frequency (in Hz)	***Link to the audio file***
16	https://drive.google.com/file/d/1LmtCkfXyRdPEaSmRIQPJAbgxrLKZHfaN/view?usp=sharing
50	https://drive.google.com/file/d/1JSCTRHl_C1oW7Mh6eI8SNb8wtwH-CNtJ/view?usp=sharing
160	https://drive.google.com/file/d/1lKgm3kviGtu1euAdT4FNGEb2r4KGC0uA/view?usp=sharing
500	https://drive.google.com/file/d/1wpRTNBI38ojYCydjzoM70OkEw1sBOUOM/view?usp=sharing

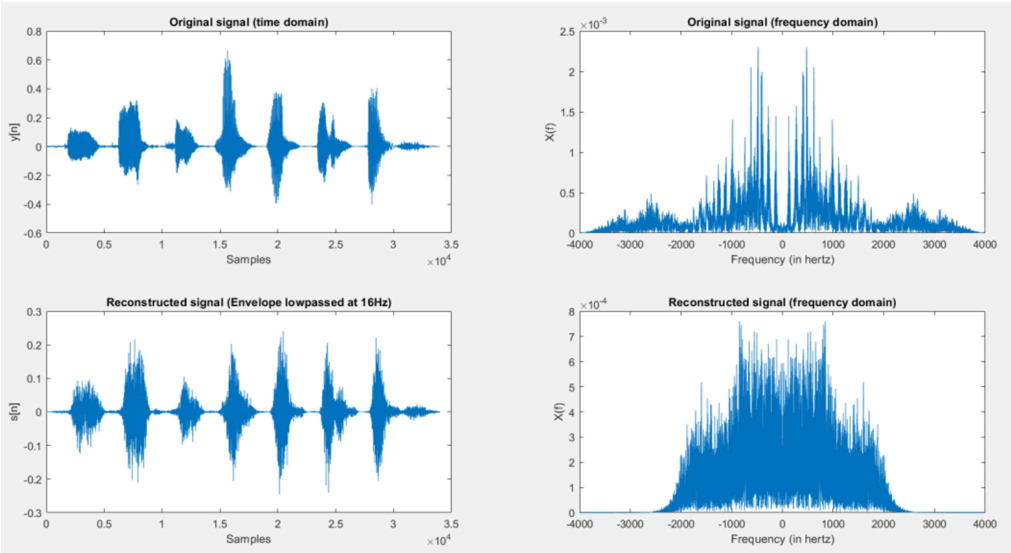

Fig. 6. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 16 Hz

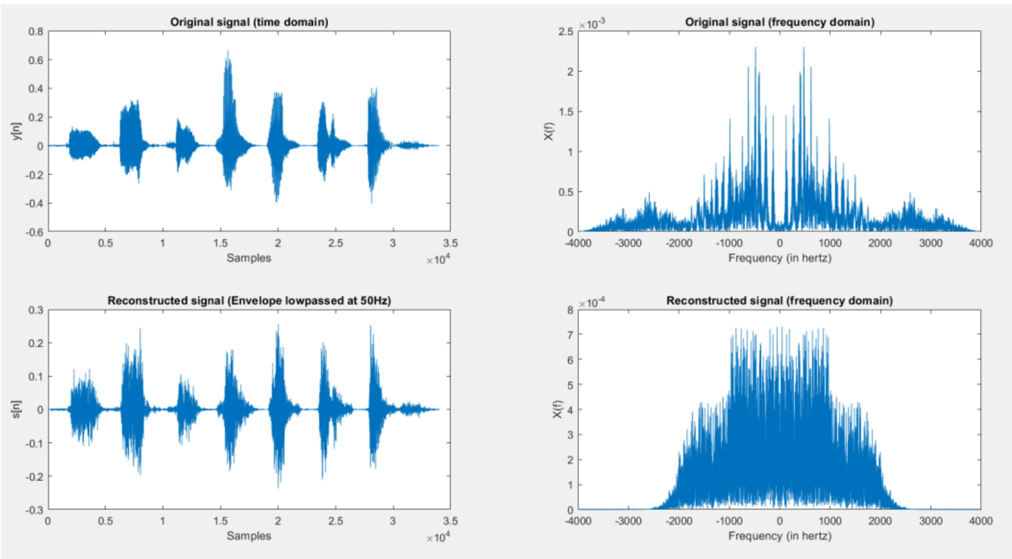

Fig. 7. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 50 Hz

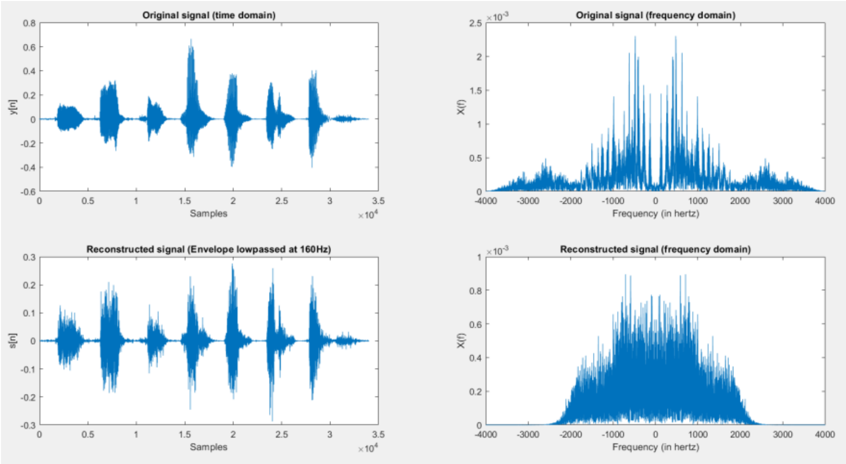

Fig. 8. ***Original V Reconstructed***: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 160 Hz

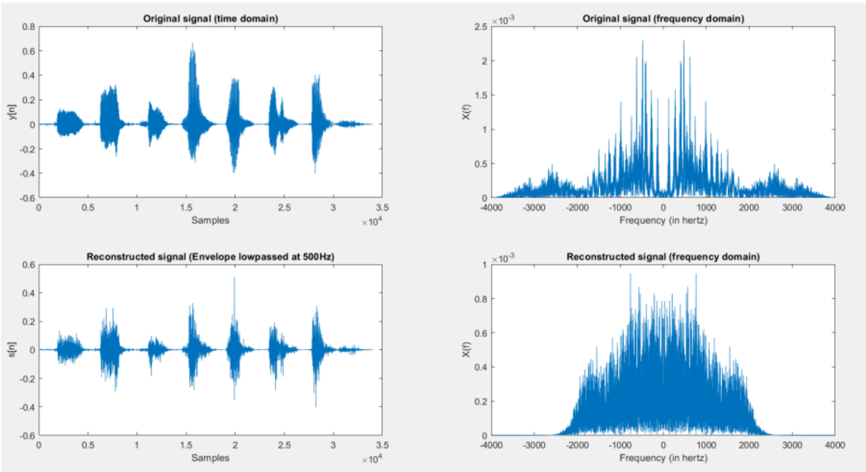

Fig. 9. ***Original V Reconstructed***: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 500 Hz

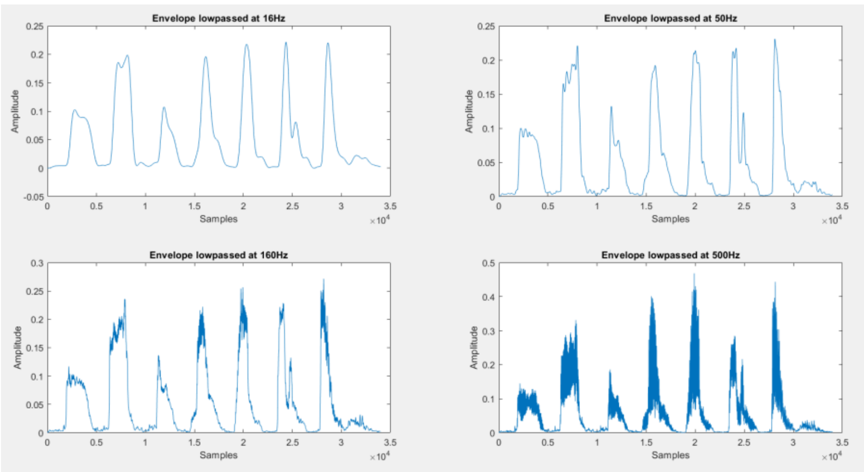

Fig. 10. ***Amplitude envelope***: For this plot, the amplitude envelope low-passed at 16 Hz, 50 Hz, 160 Hz and 500 Hz respectively, when two frequency bands are used, are shown

4.3 Using three frequency bands

In this case, the signal is using three filters:

1. Filter 1: Low-pass filter with cut-off frequency 0.8 KHz
2. Filter 2: Band-pass filter with passband 0.8 KHz – 1.5 KHz
3. Filter 3: Band-pass filter with passband 1.5 KHz – 4 KHz

The amplitude envelope is extracted and used for modulating white noise. The output file can be accessed from:

Envelope cut-off frequency (in Hz)	***Link to the audio file***
16	https://drive.google.com/file/d/1yLMVcBhiMksgD54mwmoUG_kPWJTAUkhC/view?usp=sharing
50	https://drive.google.com/file/d/1jv_79y2VfcDvQOuOrWeKQPjkxX-LWSXq/view?usp=sharing
160	https://drive.google.com/file/d/1VLWOFETASSi256YxGbvecbXChT1OfVWU/view?usp=sharing
500	https://drive.google.com/file/d/1gUvFNytIDfrC1b5Fcxp_KkeZF3vyOiD4/view?usp=sharing

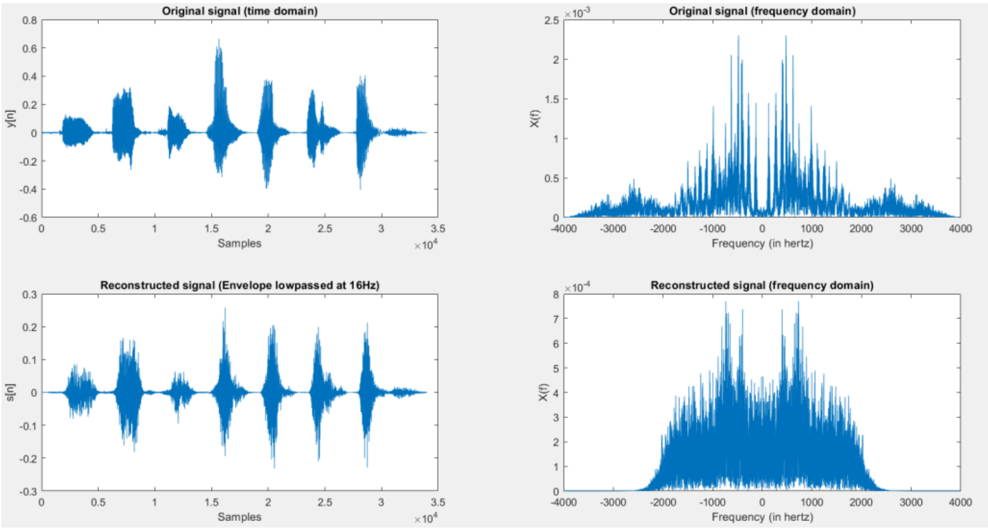

Fig. 11. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 16 Hz

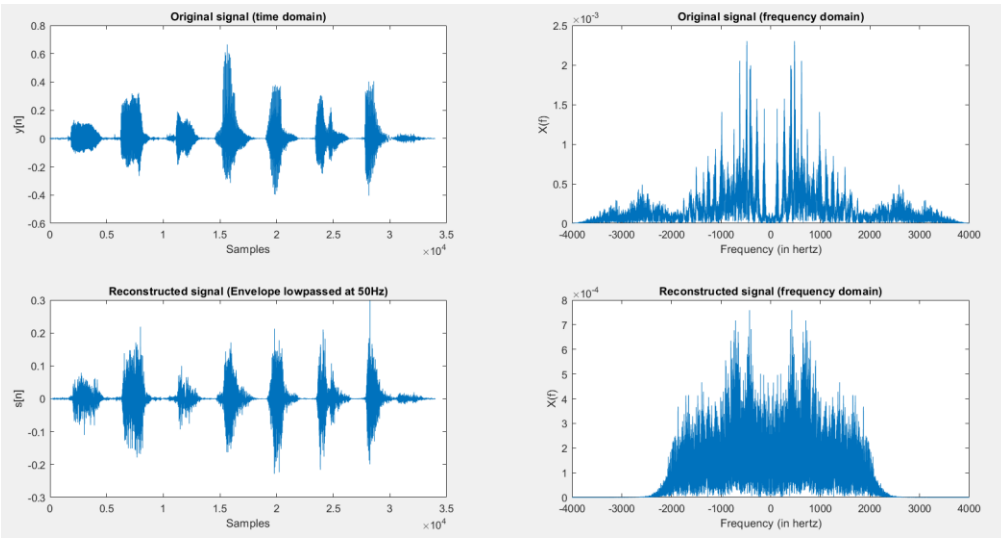

Fig. 12. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 50 Hz

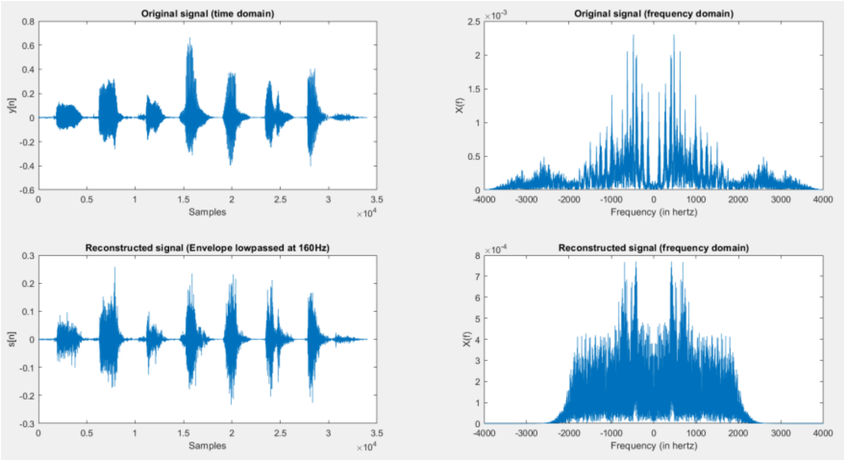

Fig. 13. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 160 Hz

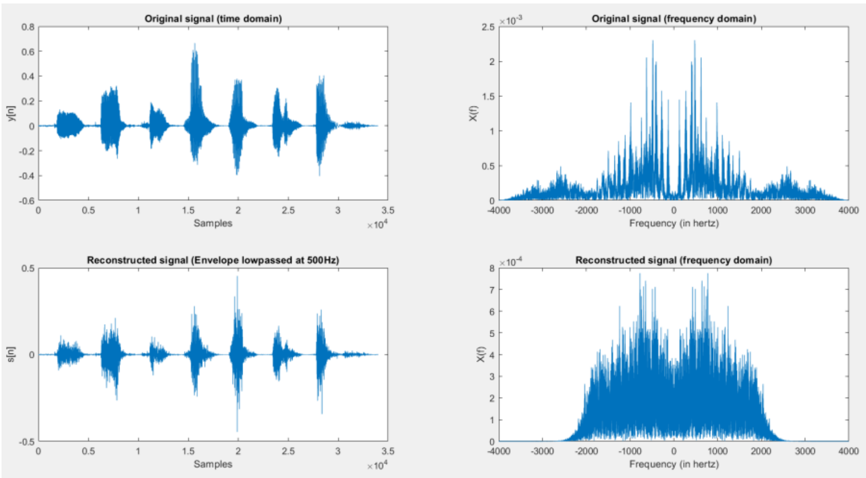

Fig. 14. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 500 Hz

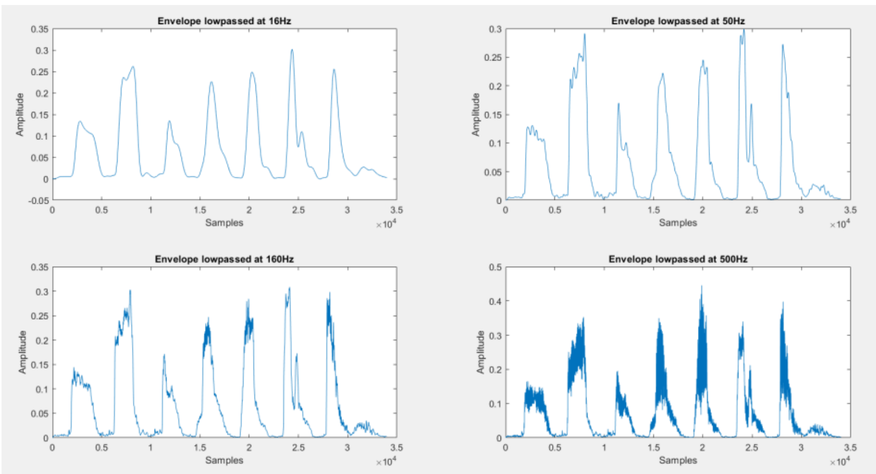

Fig. 15. *Amplitude envelope*: For this plot, the amplitude envelope low-passed at 16 Hz, 50 Hz, 160 Hz and 500 Hz respectively, when three frequency bands are used, are shown

4.4 Using four frequency bands

In this case, the signal is using four filters:

1. Filter 1: Low-pass filter with cut-off frequency 0.8 KHz
2. Filter 2: Band-pass filter with passband 0.8 KHz – 1.5 KHz
3. Filter 3: Band-pass filter with passband 1.5 KHz – 2.5 KHz
4. Filter 4: Band-pass filter with passband 2.5 KHz – 4 KHz

The amplitude envelope is extracted and used for modulating white noise. The output file can be accessed from:

Envelope cut-off frequency (in Hz)	***Link to the audio file***
16	https://drive.google.com/file/d/1_cULLbEbcA0y3Yvkk4q5yLjJ4tE9OwGG/view?usp=sharing
50	https://drive.google.com/file/d/1werdbsWNSLWsVQZMUOYWavoWGpCT4Dx7/view?usp=sharing
160	https://drive.google.com/file/d/1UK5gv-uDVekO-PPXlWEOcu4liLwEMlZT/view?usp=sharing
500	https://drive.google.com/file/d/1-mb5762bDP7_H5jsdHd5SVDuxIJTUO-J/view?usp=sharing

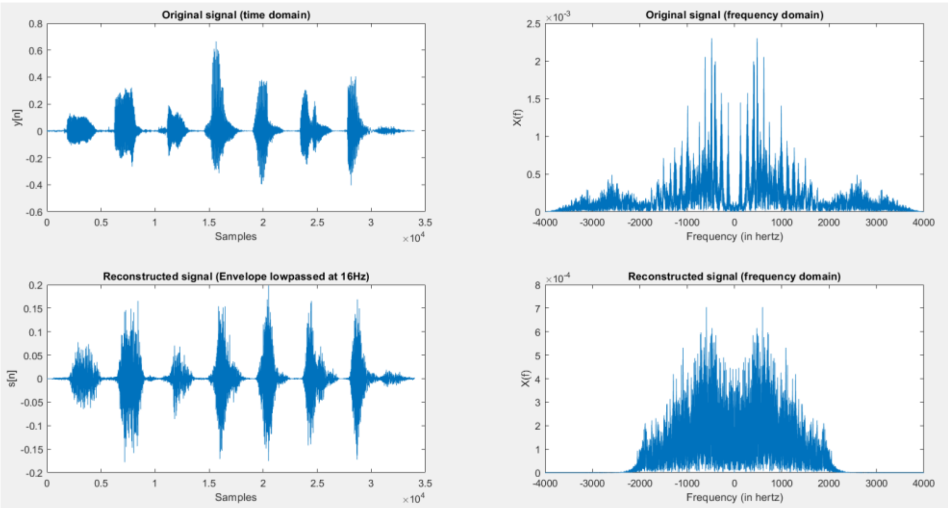

Fig. 16. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 16 Hz

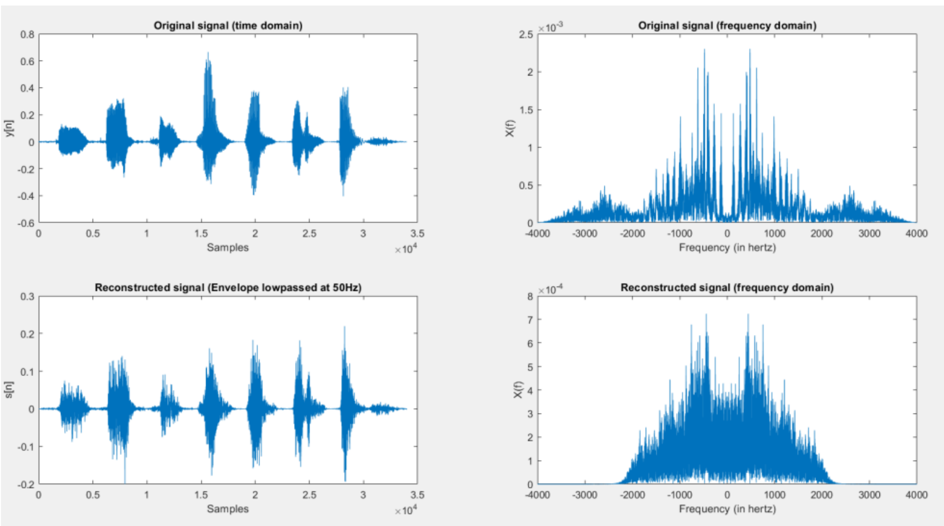

Fig. 17. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 50 Hz

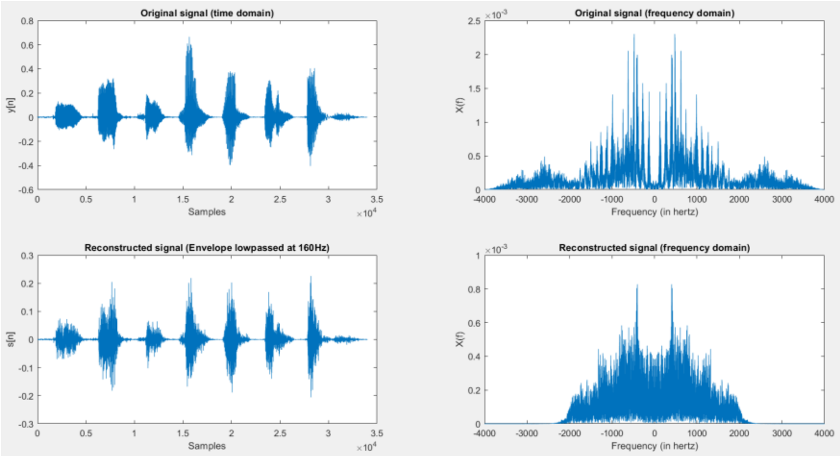

Fig. 18. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 50 Hz

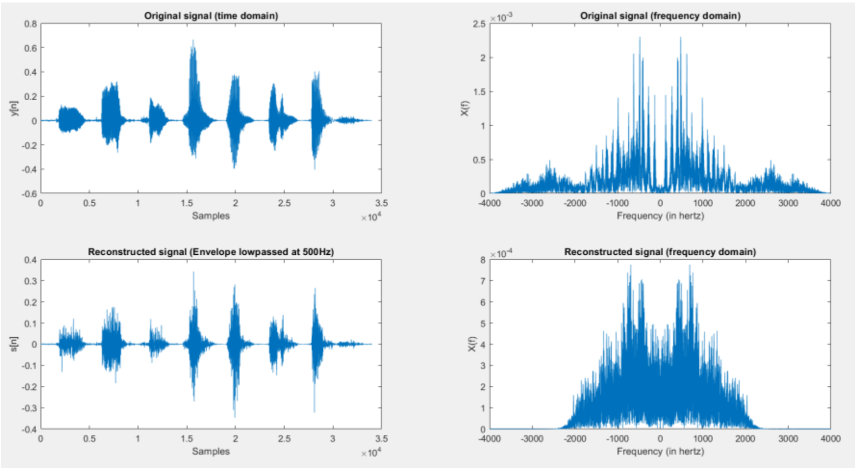

Fig. 19. *Original V Reconstructed*: For this plot, only one frequency band is used to split the original signal and the envelope is low-passed at 50 Hz

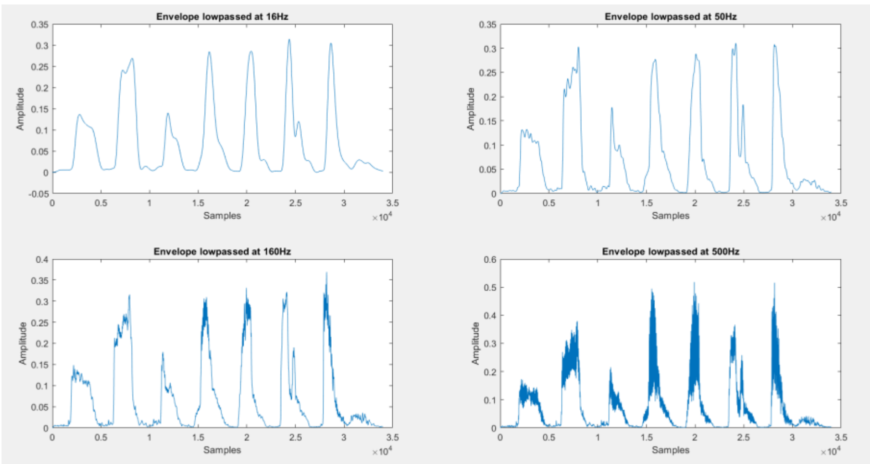

Fig. 20. *Amplitude envelope*: For this plot, the amplitude envelope low-passed at 16 Hz, 50 Hz, 160 Hz and 500 Hz respectively, when four frequency bands are used, are shown

5 Discussion

- In this experiment, the main point which we want to show is that, theoretically, to reconstruct a speech signal with 100% accuracy, we need to filter the signal for every frequency, which is practically infeasible and absolutely unnecessary. The envelope of the signal contains the information and the higher frequency components can be removed and filled with random white noise.

- The envelope extraction is done by using Hilbert transform. This helped to preserve the temporal and amplitude cues while removing the spectral detail in each band.

- The envelopes were then passed through low-pass filter with cut-off frequencies of 16 Hz, 50 Hz, 160 Hz and 500 Hz to evaluate the effect of reducing the bandwidth of temporal envelope extraction.

- The original audio signal is digitalized at 8 KHz sampling rate. The signal was then split into frequency bands depending upon the total number of filters used (which can be 1, 2, 3 or 4).

- The filters were designed with following description:

  Case 1: When only single filter is used
    - Filter 1: Low-pass filter with cut-off frequency 4 KHz

  Case 2: When two filters are used
    - Filter 1: Low-pass filter with cut-off frequency 1.5 KHz
    - Filter 2: Band-pass filter with passband 1.5 KHz – 4 KHz

  Case 3: When three filters are used
    - Filter 1: Low-pass filter with cut-off frequency 800 Hz
    - Filter 2: Band-pass filter with passband 0.8 KHz – 1.5 KHz
    - Filter 3: Band-pass filter with passband 1.5 KHz – 4 KHz

  Case 4: When four filters are used
    - Filter 1: Low-pass filter with cut-off frequency 800 Hz
    - Filter 2: Band-pass filter with passband 0.8 KHz – 1.5 KHz
    - Filter 3: Band-pass filter with passband 1.5 KHz – 2.5 KHz
    - Filter 4: Band-pass filter with passband 2.5 KHz – 4 KHz

- It was observed that the clarity of the output signal increased with an increase in the number of frequency bands used and an increase in the envelope cut-off frequency.

- Fig. 1, 2, 3, and 4 show that the reconstructed signal's frequency response is entirely different from that of the original signal. However, as the frequency bands are increased, we can see that this response moves closer and closer towards the original frequency response. See Fig. 6 – 9, 11 – 14, 16 – 19.

- For a fixed number of frequency bands, the quality of the envelope improves significantly with an increase in envelope cut-off frequency.

- For a fixed envelope cut-off frequency, the quality of the envelope also improves with the increase in frequency bands. See Fig. 5, 10, 15, and 20.

- Though the clarity of the output signal increases with the increase in frequency bands and envelope cut-off frequency, for a person listening to the output speech for the first time, it is still difficult to completely identify the words.

6 MATLAB

```matlab
% Author: Utkarsh Patel
% Experiment 6

[y, Fs] = audioread('audio.wav'); % reading the audio signal

N = length(y); % length of the signal

bpf = [10 4000 0 0 0; 10 1500 4000 0 0; 10 800 1500 4000 0; 10 800 1500 2500 4000];
% matrix containing frequency required for constructing filters

lpf = [16 50 160 500]; % envelope cut-off frequency

nbpf = 4; % change this to alter number of frequency bands

recon = containers.Map({'1', '2', '3', '4'}, {[], [], [], []});
% map object for storing envelopes for all the four envelopes cut-off frequency

for r = 1: 4 % choosing the envelope cut-off freq
    x = 0;   % variable to store envelope
    n=randn(1,N); % white noise
    s=zeros(1,N); % variable to store reconstructed signal
    for i = 1: nbpf
        [b, a] = butter(4, [bpf(nbpf, i) / Fs, bpf(nbpf, i + 1) / Fs], 'bandpass');
        yfilt=filter(b,a,y); % filtering signal for given passband
        nfilt=filter(b,a,n); % filtering noise for given passband
        env = abs(hilbert(yfilt)); % extracting envelope
        [b1, a1] = butter(4, lpf(r) /Fs, 'low');
        env = filter(b1, a1, env); % filtering envelope
        x = x + env;               % summing envelope
        env=transpose(env);
        s=s+(env.*nfilt);          % modulating noise with envelope
    end

    recon(string(r)) = [recon(string(r)); x]; % storing envelope to map object
    [b, a] = butter(4, 4000 / Fs, 'low');
    s = filter(b, a, s);  % low-passing reconstructed signal


    figure;

    subplot(2, 2, 1);
    plot(y);
    title('Original signal (time domain)');
    xlabel('Samples');
    ylabel('y[n]');

    subplot(2, 2, 2);
    dF = Fs/N;
    f = -Fs/2:dF:Fs/2-dF;
    Y = fftshift(fft(y));
    plot(f,abs(Y)/N);
    title('Original signal (frequency domain)');
    xlabel('Frequency (in hertz)');
    ylabel('X(f)');

    subplot(2, 2, 3);
    plot(s);
    title('Reconstructed signal (Envelope lowpassed at ' + string(lpf(r)) + 'Hz)');
    xlabel('Samples');
    ylabel('s[n]');

    subplot(2, 2, 4);
    dF = Fs/N;
    f = -Fs/2:dF:Fs/2-dF;
    S = fftshift(fft(s));
    plot(f,abs(S)/N);
    title('Reconstructed signal (frequency domain)');
    xlabel('Frequency (in hertz)');
    ylabel('X(f)');
end

figure;

for r = 1: 4
    subplot(2, 2, r);
    plot(recon(string(r)));
    title('Envelope lowpassed at ' + string(lpf(r)) + 'Hz');
    xlabel('Samples')
    ylabel('Amplitude')
end
```