Audio Filtering and Clap Detection

Vibha Rao

1 Audio Filtering

This project involves filtering a noisy speech signal sampled at 48,000 Hz using digital bandpass filters. Gaussian white noise was artificially added to the clean speech signal to simulate a noisy environment. Both Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) filters were implemented to isolate the desired frequency components of the speech signal. Time and frequency domain analyzes were performed before and after filtering to visualize the effect of filtering.

1.1 Characteristics of Speech Signal

Human speech signals typically occupy a frequency range of approximately 300 Hz to 3400 Hz, which contains most of the intelligible information. Normalization is applied to ensure that the signal amplitude is scaled to a range between -1 and 1.

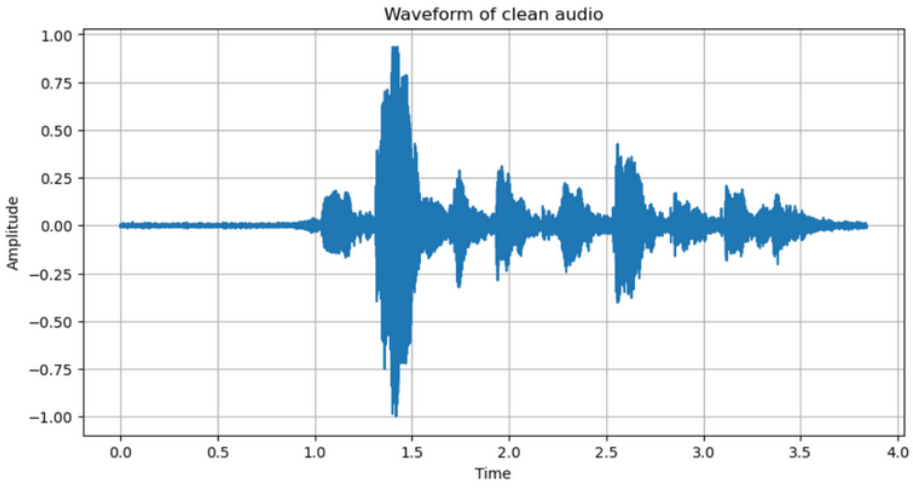


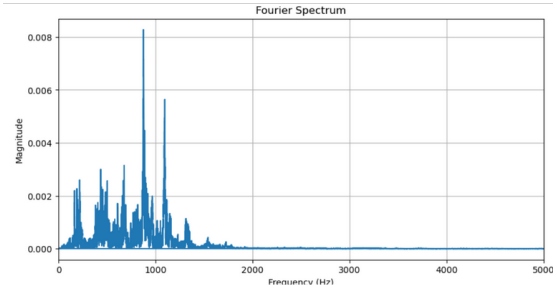
Figure 1: Caption describing the figure.

FFT and Spectrogram Observations

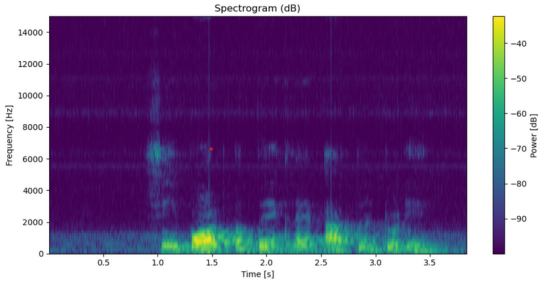
Spectrogram is a time-frequency representation of a signal, created by computing the Short-Time Fourier Transform (STFT) over successive time windows. It shows how the frequency content

of a signal evolves over time

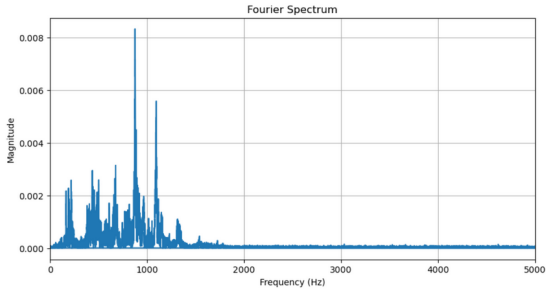
- The FFT of clean speech signal shows concentrated energy in the 300–3400 Hz range
- The spectrogram of clean signal exhibits distinct vertical patterns corresponding to phonemes and formants.
- The noisy signal appears smeared and dense across all frequencies.



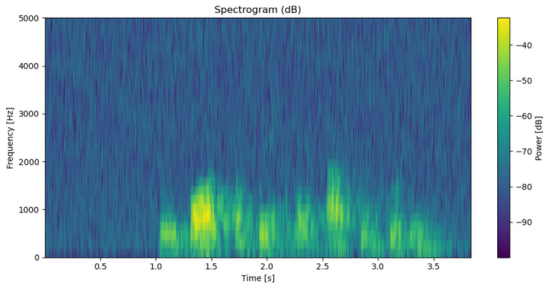
(a) FFT of Clean Signal



(b) Spectrogram of Clean Signal



(c) FFT of Noisy Signal



(d) Spectrogram of Noisy Signal

Figure 2: Comparison of original, noisy, and filtered speech signals.

1.2 FIR Filter

A Finite Impulse Response (FIR) filter is a type of digital filter whose output depends solely on a finite number of past and present input values.

- They are stable with linear phase shift.
- FIR filtering is implemented using the discrete convolution.
- Impulse response of FIR Filter is a windowed sinc function
- Computationally expensive

The output $y[n]$ of an FIR filter of order N is

$$y[n] = \sum_{k=0}^{N} h[k] \cdot x[n-k] \tag{1}$$

- $x[n]$ is the input signal,
- $h[k]$ are the filter coefficients (impulse response),
- $y[n]$ is the output signal,
- N is the filter order.

A bandpass fir filter of size 801 with cutoff from 300 Hz to 3400 Hz is designed using firwin() scipy function. The frequency and phase response are shown in Figure 3. We can clearly observe a sharp cutoff and linear phase shift.

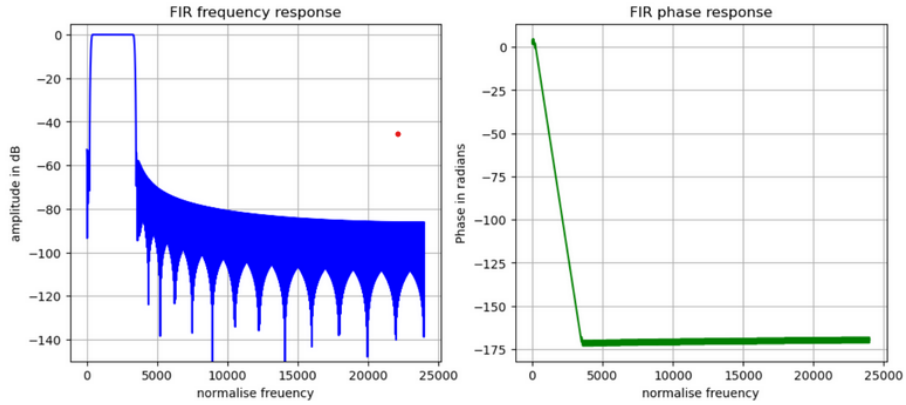


Figure 3: Caption describing the figure.

1.3 IIR Filter

An Infinite Impulse Response (IIR) filter is a digital filter that utilizes feedback, meaning its output depends on both current and past input values as well as past output values.

- They cause a non linear phase shift, can cause distortion
- Can face stability issues
- Computationally efficient.
- Used for real time filtering

$$y[n] = \sum_{k=0}^{M} b_k \, x[n-k] - \sum_{j=1}^{N} a_j \, y[n-j] \tag{2}$$

- $x[n]$ is the input signal,
- $y[n]$ is the output signal,
- b_k are the feedforward coefficients,
- a_j are the feedback coefficients,
- M and N define the order of the feedforward and feedback parts respectively.

IIR digital filters are often designed by transforming analog filters into the digital domain. Consider the low RC circuit in the figure, the transfer function in s domain is given by

$$H(s) = \frac{1}{1 + RCs} \tag{3}$$

To convert this analog filter into a digital IIR filter, the bilinear transform is applied:

$$H(z) = \frac{2}{T}\frac{1 - z^{-1}}{1 + z^{-1}} \tag{4}$$

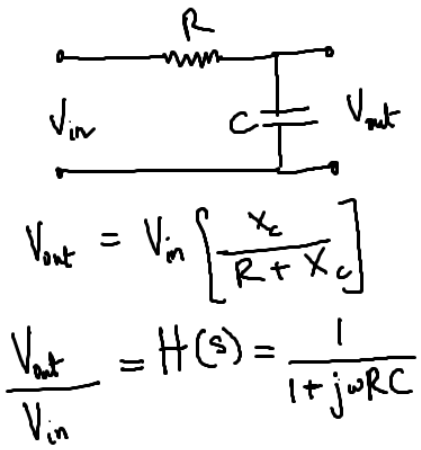


Figure 4: Response of IIR Filter.

Thee filter characteristics for a 4th order IIR bandpass filter with cutoff frequency from 300 Hz to 3400 Hz are shown in the Figure 4.

1.4 Output

The waveform plots of the FIR and IIR filtered signals show clear noise reduction compared to the noisy input. **FIR-filtered spectrogram** and **IIR-filtered spectrogram** reveals a clear restoration of the speech band (300–3400 Hz), with significant suppression of out-of-band noise.

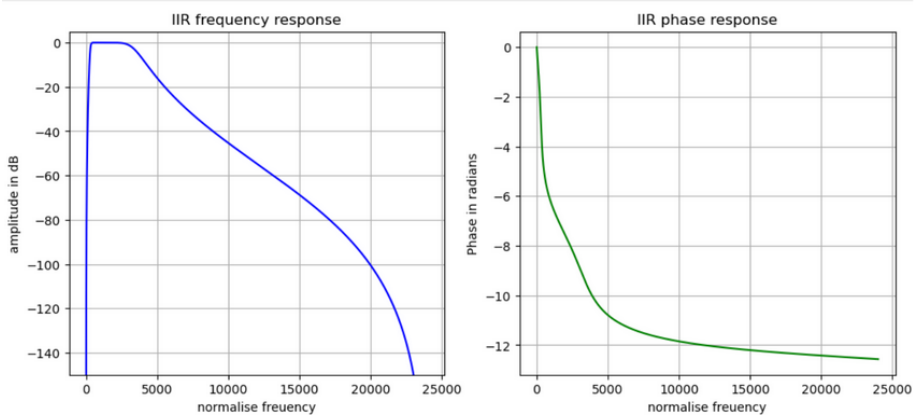
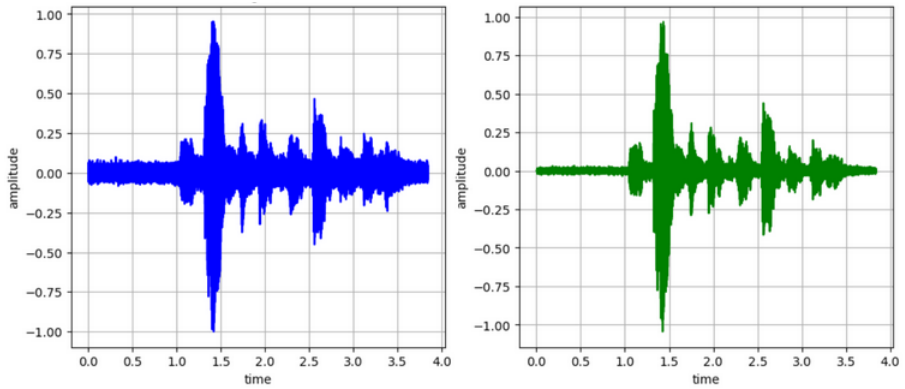
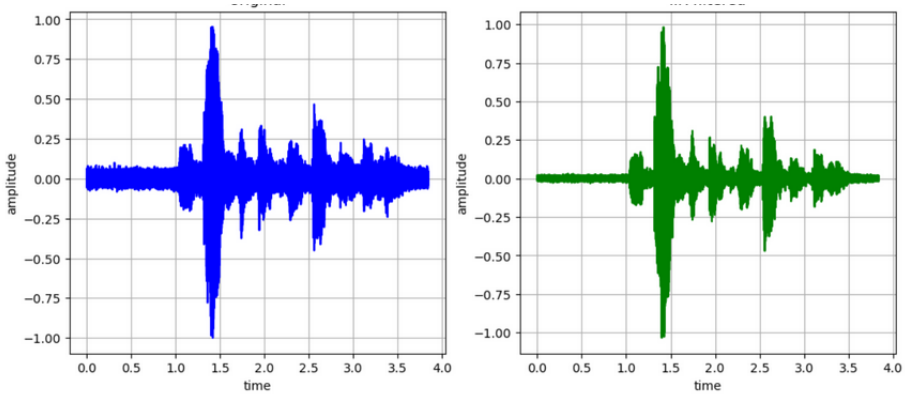


Figure 5: Response of IIR Filter.

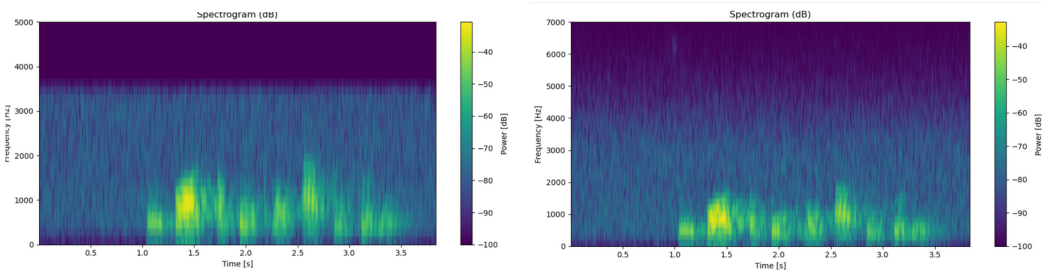


(a) Noisy vs FIR Filtered Waveform



(b) Noisy vs IIR Filtered Waveform

Figure 6: Comparison of FIR and IIR filtered signals.



(a) Spectrogram of FIR Filtered signal

(b) Spectrogram of IIR Filtered signal

Figure 7: Spectrgoram of filtered sigals

2 Clap Detection

This project focuses on detecting events like a clap within an audio signal. The test audio contains a mix of speech, clapping sounds, and a spoon clanging. Frequency and time domain characteristics are visualized and used to analyzed to detect sharp tranisent sounds.

2.1 Methodology

To achieve this, the signal is analyzed using the following methods:

- Short-Time Energy (STE): captures sudden bursts in energy typical of claps.
- Band Energy Ratios: used to distinguish claps from other transients based on their frequency content.

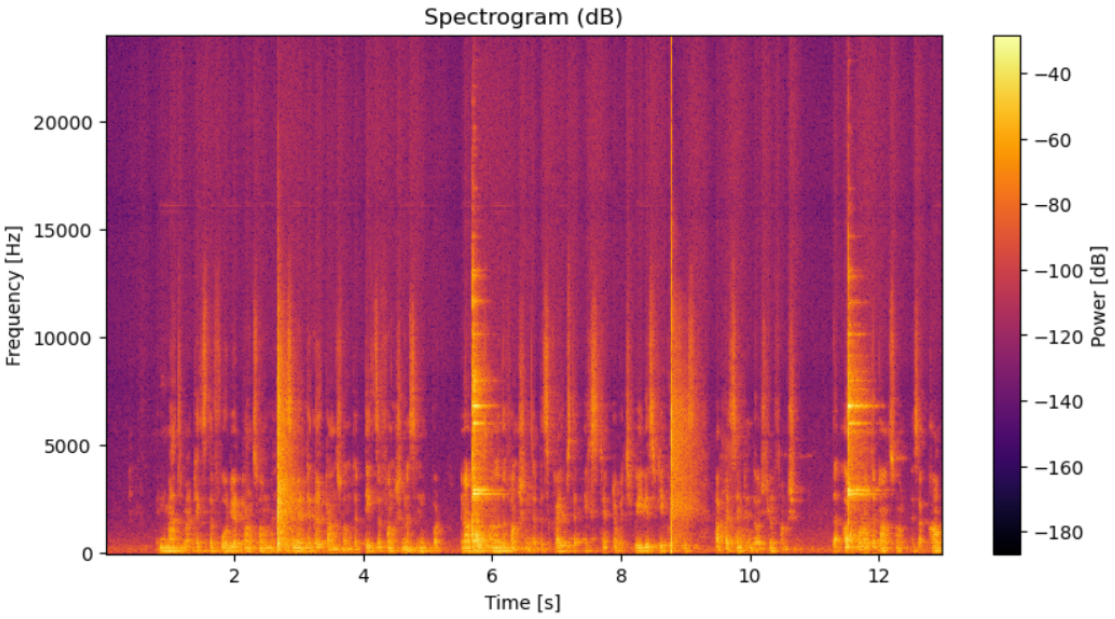


Figure 8: Spectrogram of signal. Transients appear as long bright spikes across frequencies.

2.2 Spectrogram Analysis of Transients

Transients such as claps or metallic clangs appear as short-duration, high-energy bursts across a wide frequency range. From the spectrogram we can observe

- Claps produce a sharp vertical stripe in the spectrogram, indicating a sudden, broadband energy burst. The energy dies quickly.

- Spoon clanging also results in high-energy transients. The energy distribution is longer than a clap due to ringing.

These can be observed from the spectrogram plot. Claps occur at 3 second and 9 second mark. Spoon sound occurs at 6 second and 12 second mark.

2.3 Detection

To detect transient events we use signal features computed over short overlapping frames.

Short Time Energy

Short-Time Energy measures the signal's power within a short window of time. The signal is divided into small frames of 10ms with a 5ms overlap. Each frame is windowed using a hanning window and the squared amplitude values in each frame are summed. A Hann window is used over a rectangular window to prevenet sharp edges and spectral leakage. Transient sounds produces high, narrow peaks in the STE curve, while speech shows smoother, longer energy variations as in Figure 8.

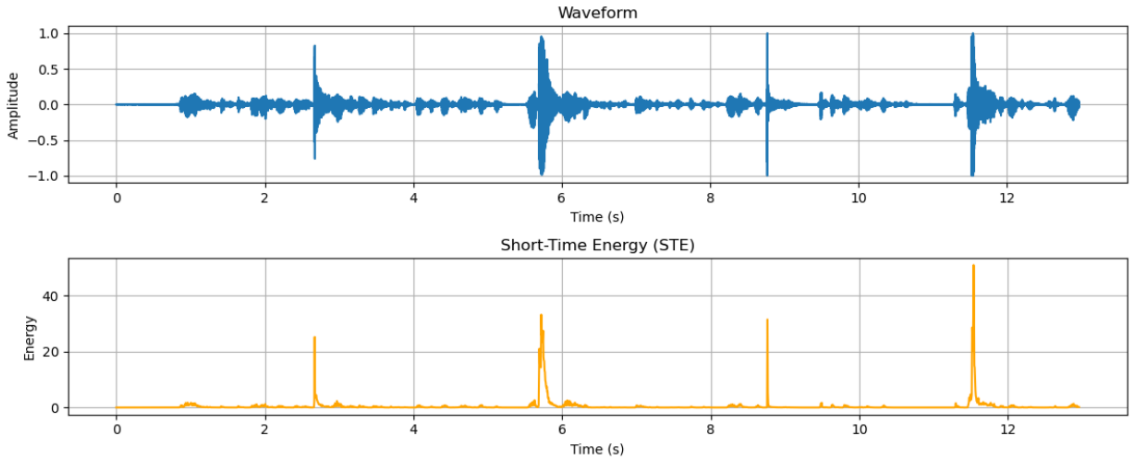


Figure 9: Short Time Energy

Band Energy Ratio

It is clear from spectrogram that spoon and clap sounds cause high energy at high frequency bands. Therefore by calculating the ratio of energies at low frequency (500-400) and high frequency (15000-20000Hz) bands over time will help detect instances of transients. Figure 9 shows at transients the BER is high. A threshold of 3.38 is set for detection of tranisents.

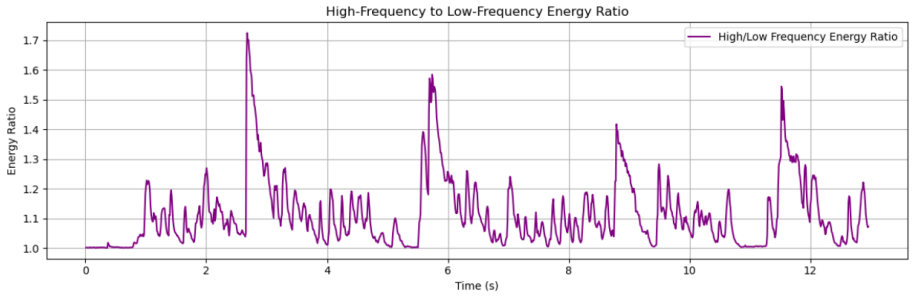
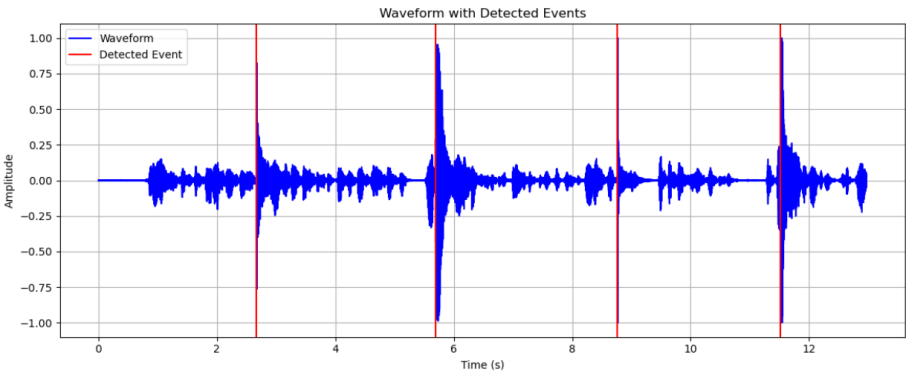
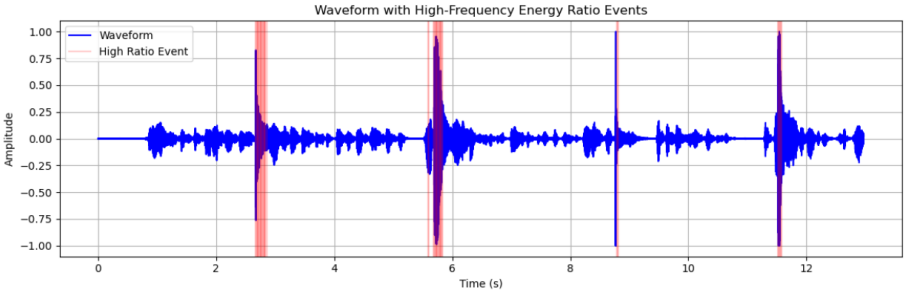


Figure 10: Band Energy Ratio



(a) Transient detection using STE



(b) Transient Detection using BER

3 Conclusion

This project explored denoising speech signals using digital filters and detecting transient acoustic sounds. Through this I learnt about design and implementation of FIR and IIR filters, filter characteristics and parameters. Along with this basic frequency, time domain analysis techniques like fft, spectrogram, short time energy, zero crossing rate were analyzed.

4 Future Work

- Learn how to compare effectiveness of FIR vs IIR Filter.
- Experiment with filter order to understand phase distortion
- For Clap Detection features could be combined to come up with a generic metric that can be applied to any signal.
- Thresholding is done manually, learn about adaptive thresholds.
- Learn math and implementation behind algorithms like STFT, FFT etc.