A Project Report

On

SPECTRAL AND TEMPORAL ANALYSIS OF VOWEL AND CONSONANT SOUNDS USING DIGITAL SIGNAL PROCESSING

A Report on Summer Internship

(10^{th} June – 18^{th} July 2025)

Submitted by

Harun Ansari

Roll No- 10200323025

Department of Electronics and Communication Engineering

Kalyani Government Engineering College

Under the Supervision of

Dr. Jishan Mehedi

Head of the department of Electronics and Communication Engineering Department in Jalpaiguri Government Engineering College

Acknowledgements

I would like to express my thanks to all the faculty members specially Dr. Jishan Mehedi sir , Jalpaiguri Govt. Engg. College for instructing us during the whole Semester for motivating constantly, providing this wonderful innovative project, preparation of this seminar & report, acquainting us with the required details, & providing us with this wonderful opportunity. To the best of my knowledge, the results embodied in this report, are original in nature and worthy of incorporation in the present version of the report.

Name-Harun Ansari **Roll no-10200323025** **Signature**

Harun Ansari

Undertaking

I, Harun Ansari (Roll No:10200323025) student of B.Tech in Electronics & Communication Engineering of Kalyani Government Engineering College here by declare that Internship Project report on "**Spectral and Temporal Analysis of Vowel and Consonant Sounds Using Digital Signal Processing**" is our own contribution. I also declare that I have not indulged in any form of plagiarism to carry out this Project and I am only responsible for the words written in the internship project report. The work or idea of other authors which are utilized in this report are properly acknowledged and mentioned in the References. I undertake total responsibility if any trace of plagiarism is found at any later of the stage.

Name – Harun Ansari **Roll no-10200323025** **Signature**

- **Harun Ansari**

Jalpaiguri Government Engineering College

A Govt. Autonomous Engg. College

(Affiliated to Maulana Abul Kalam Azad University of

Technology) Jalpaiguri – 735102, WB, India

<u>Certificate of Approval</u>

This is to certify that the internship project report on **“Spectral and Temporal Analysis of Vowel and Consonant Sounds Using Digital Signal Processing”** is a record of Bonafide work, carried out by “Harun Ansari” of “Kalyani Government Engineering College ” under my guidance and supervision.

In my opinion, the report in its present form is in conformity as specified by **Jalpaiguri Government Engineering College** and as per regulations of the Maulana Abul Kalam Azad University of Technology (MAKAUT). To the best of my knowledge the results presented here are original in nature and worthy of incorporation in internship project report. for the B.Tech. program in Electronics & Communication in the academic year 2021-2025.

Signature of	**Signature of**
Supervisor	Head, Dept of ECE,JGEC

Contents

SL No.	Topic	Page No.
1.	Abstract	06
2.	Introduction	07
3.	Literature Review	08
4.	Problem Formulation	09
5.	Technique	09
6.	Step	10
7.	Solution	11
8.	Result And Analysis	12-14
9.	Conclusion	14
10.	References	15
11.	Appendix	16-17

1.Abstract

This project drives into a comparative analysis of vowel and consonant sounds through digital signal processing (DSP) techniques. Speech is made up of various phonetic elements, and vowels and consonants stand out with their distinct acoustic and spectral properties. The main goal here is to differentiate between a vowel sound, like “A” and a consonant sound, such as “H” by examining their characteristics in both the time and frequency domains. To carry out this analysis, we used two recorded audio files one capturing the vowel and the other the consonant. MATLAB is our tool for processing the signals, which included tasks like plotting waveforms, performing Fast Fourier Transform (FFT), and calculating the spectral centroid. When we looked at the time-domain analysis the vowel signal displayed smooth, periodic waveforms that indicated a steady vibration of the vocal folds. In contrast, the consonant signal had sharp, irregular patterns, showcasing the quick articulatory movements involved. The frequency domain analysis with FFT showed that vowel energy tends to cluster in the low frequency range, while consonants pack a punch with their high-frequency components. Additionally, we calculated the spectral centroid as a way to quantify the distribution of high frequency energy. We found that the centroid for the consonant was not probably higher than that of the vowel, which confirms that consonants demand more effort in articulation and display greater spectral complexity. This project effectively illustrates how we can use signal processing tools to tell apart vowels and consonants and the mathematical analysis of their spectral features sheds light on the mechanisms behind speech production. This approach has the potential to be expanded into areas like speech classification phoneme recognition, or even applications in speech therapy.

2.Introduction

Speech sounds are generally divided into vowels and consonants each having unique properties. Vowels tend to have sustained periodic waveforms with well-defined resonant peaks referred to as formants, whereas consonants tend to have transient noise like bursts or sudden changes in the spectrum. Through an analysis of both spectral and temporal domains, we can reveal the fine markers that distinguish these phonetic units and gain a deeper appreciation of their contribution to human communication.

Examination of these a properties is of vital importance to technologies like automatic speech recognition, speaker identification, hearing aid design, and linguistic studies. Spectral information such as formant frequencies and spectral centroids encode the resonant nature of sounds, while temporal cues like segment durations and voice onset times emphasize dynamic properties that are critical to discriminating plosives, fricatives.

Although numerous studies have been conducted on either spectral or temporal characteristics alone few have investigated both analyses together within a single digital signal processing stream. Most current solutions base their methods on a proprietary toolbox or are limited to particular phoneme categories with no systematic comparison across a wide range of vowels and consonants. This project fills that void by using a script based MATLAB that extracts and displays spectral and temporal features together.

3.Literature Review

The exploration of speech signal processing has been a key focus in linguistics, electrical engineering, and computer science for many years. Researchers have delved into the acoustic characteristics of speech sounds, particularly vowels and consonants, employing a variety of analytical methods such as time-domain, frequency-domain, and machine learning techniques. A landmark study in this field was conducted by Rabiner and Schafer (2010), where they outlined how speech can be modelled and analysed through digital signal processing. They pointed out that vowels, because they are voiced, have rich harmonic content and energy that tends to cluster in the lower frequency range. In contrast, consonants are characterized by quick transitions, high zero-crossing rates, and often include unvoiced elements, especially in sounds like fricatives and plosives. Numerous studies have shown that temporal features, such as short-time energy and zero-crossing rate (ZCR), are quite effective for segmenting and classifying different phonemes. For example, consonants usually exhibit a higher ZCR due to their noisier characteristics, while vowels display a lower ZCR and greater overall signal energy. In the spectral realm, the idea of the spectral centroid has been instrumental in differentiating between bright (high-frequency dominant) and dull (low-frequency dominant) sounds. Research indicates that consonant sounds generally have a higher spectral centroid compared to vowels, thanks to the presence of high-frequency noise or transients. Additionally, recent advancements in automatic speech recognition (ASR) systems and phoneme classification have leveraged these essential DSP-based features before integrating them into machine learning models. While modern deep learning techniques can automatically learn features, traditional handcrafted features like energy, ZCR, and spectral centroid continue to play a significant role in applications such as speech segmentation, emotion detection, and speaker recognition.

Even though there's been a ton of research into speech signal processing, we still need to find better ways to simplify and visualize the differences between vowels and consonants, especially for educational, linguistic, and experimental uses. This project is all about making a contribution in that area by using MATLAB-based DSP techniques to pull out and compare the spectral and temporal features of vowel and consonant sounds.

4.Problem Formulation

Speech is one of the most natural and common ways we humans communicate. It involves various types of sounds, primarily vowels and consonants, each with its own unique characteristics in how they're produced and perceived. Vowels tend to be voiced, periodic, and full of energy, while consonants can be either voiced or unvoiced, often exhibiting transient or noisy traits. From a signal processing perspective, these distinctions manifest in both temporal (time-domain) and spectral (frequency-domain) features. In automated speech processing systems like speech recognition, phoneme classification or voice-controlled interfaces differentiating between vowel and consonant sounds based on their acoustic signals can be quite challenging. Factors like background noise, speaker variations, and rapid speech changes can negatively impact performance if the system is not finely tuned to recognize these differences. To enhance speech analysis systems, we need to explore straightforward yet effective signal features that can help distinguish vowels from consonants. Although there are numerous machine learning models out there, they often rely on handcrafted or learned features grounded in fundamental digital signal processing concepts.

5.Technique

- **Tools and Technologies:**

1.Software: MATLAB R2024a

2.Input: .mp4a audio recordings of vowel and consonant sounds

3.Sampling Frequency: 48KHz

6.Step

Audio input -----> Time-Domain Analysis ----->Frequency-Domain Analysis -----> Feature Extraction -----> Comparison and Plotting

- **Audio input**:

  Using audio read () to load vowel and consonant recordings.

- **Time-Domain Analysis**:

  Plotting the waveform of each signal.

- **Frequency-Domain Analysis**:

  Applying FFT to visualize spectral content.

- **Feature Extraction**:

  Energy

  Zero-Crossing Rate (ZCR)

  Spectral Centroid

- **Comparison and Plotting:**

  Vowel vs Consonant feature comparison using graphs and data.

7.Solution

- **Spectral Centroid (Main Metric of "Effort"):**

$$SC = \frac{\sum_{k=0}^{N-1} f(k) \cdot |X(k)|}{\sum_{k=0}^{N-1} |X(k)|}$$

X(k) = magnitude of the FFT at frequency bin k

f(k) = frequency corresponding to bin k

SC = spectral centroid

- **Zero-Crossing Rate (ZCR):**

$$ZCR = \frac{1}{2N} \sum_{n=1}^{N} |\text{sgn}(x[n]) - \text{sgn}(x[n-1])|$$

$$\text{Sgn}(x[\text{n}]) = \begin{cases} +1, & x[n] > 0 \\ 0, & x[n] = 0 \\ -1, & x[n] < 0 \end{cases}$$

- **Fast Fourier Transform (FFT):**

$$X(k) = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi kn}{N}}$$

$X[n]$= time domain signal

$X[k]$= frequency domain representation

N = total number of samples

8. **Result And Analysis**

- **Time-Domain Analysis:**

The waveform of both the vowel and consonant sounds was observed in the time domain**:**

- **Vowel (A):**

  In this graph it Shows strong, horizontal bands especially after 1.0s. This indicates strong periodicity.

- **Consonant (H):**

  In this graph it shows sporadic with vertical patterns indicating temporary noise over frequency.

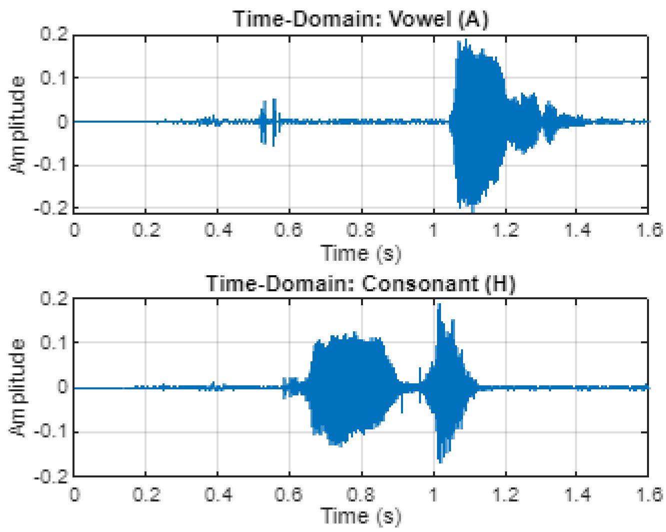

Fig1: Time Domain of Vowel and Consonant

> ➢ **<u>Frequency-Domain (FFT) Analysis:</u>**

The FFT magnitude spectrum for both signals was plotted:

- **<u>Vowel (A)</u>**:

  1.Most of the energy is concentrated at lower frequencies .

  2.There is a drop in magnitude after initial vertex, then continue with same mid frequencies and rise again before the end.

  3.Indicates a strong presence of low frequency harmonics.

- **<u>Consonant(H)</u>**:

  1.Magnitude is spread out over frequencies .

  2.There is an initial peak and it shows higher energy across mid to high frequencies compared to the vowel .

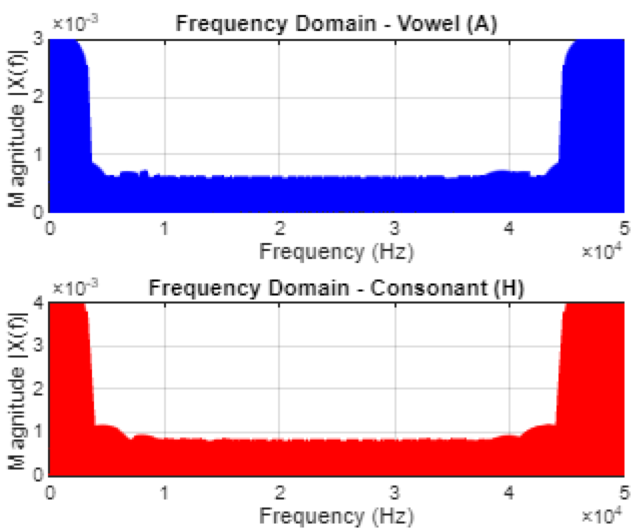

Fig2: Frequency Domain of Vowel and Consonant

- **<u>Spectral Centroid (Mathematical Effort Measure)</u>**

The spectral centroid was calculated to quantify high-frequency weight of each sound:

<u>Vowel (A):</u> 2906.18 Hz

<u>Consonant (H):</u> 4629.90 Hz

Spectral centroid of consonants are higher than vowel. It mathematically confirms that consonants need more high-frequency components.

9.Conclusion

In this project, we analysed vowel and consonant sounds using digital signal processing techniques. These included time-domain plotting, frequency-domain (FFT) analysis, and spectral centroid computation.

1. Vowel signals are smoother and more regular. They also have energy focused in the low-frequency range.
2. Consonant signals are more sudden, less regular, and have energy spread across higher frequency areas.
3. The spectral centroid for consonants was much higher than that for vowels. This shows that consonants involve more articulatory effort and contain more high-frequency content.
4. Frequency-domain and centroid analysis were effective tools for telling apart vowel and consonant pronunciations.

The analysis shows that we can easily tell apart vowel and consonant sounds by looking at their spectral properties. Plus signal processing techniques like FFT and spectral centroid give us a solid mathematical foundation for making these comparisons.

10. References

1. Tarun Kumar Rawat (2015), Digital Signal Processing, Oxford Publications.
2. Boersma, P., & Weenink, D. (2023).
3. Proakis, J. G., & Manolakis, D. G. (2007). Digital Signal Processing: Principles, Algorithms, and Applications. Pearson.
4. Chen, C. T. (2009). Analog and Digital Signal Processing. Oxford University Press.
5. Stevens, K. N. (2000).Acoustic Phonetics. MIT Press.
6. Mitra, S. K. (2011). Digital Signal Processing: A Computer-Based Approach. McGraw-Hill.
7. Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1999).
8. Wikipedia Contributors. https://en.wikipedia.org/wiki/Spectral_centroid
9. Ifeachor, E. C., & Jervis, B. W. (2002). Digital Signal Processing: A Practical Approach (2nd ed.). Pearson.
10. MathWorks Inc. (2024). https://www.mathworks.com/help/signal/
11. Huang, X., Acero, A., & Hon, H. W. (2001).

11. Appendix

- **MATLAB Code For Analysis:**

```
% === DSP Project: Vowel vs Consonant Comparison ===
% Vowel = A.m4a, Consonant = H.m4a

% Step 1: Read audio
[vowel, FsV] = audioread('A.m4a');
[consonant, FsC] = audioread('H.m4a');

vowel = vowel(:,1); consonant = consonant(:,1);

% Step 2: Trim to same length
N = min(length(vowel), length(consonant));
vowel = vowel(1:N); consonant = consonant(1:N);
t = (0:N-1)/FsV;

% Step 3: Play audio
disp('Playing Vowel (A)...'); sound(vowel, FsV); pause(length(vowel)/FsV + 1);
disp('Playing Consonant (H)...'); sound(consonant, FsC); pause(length(consonant)/FsC
+ 1);

% Step 4: Time-Domain Plot
figure;
subplot(2,1,1);
plot(t, vowel); title('Time-Domain: Vowel (A)'); xlabel('Time (s)');
ylabel('Amplitude'); grid on;
subplot(2,1,2);
plot(t, consonant); title('Time-Domain: Consonant (H)'); xlabel('Time (s)');
ylabel('Amplitude'); grid on;

% Step 5: Frequency-Domain (FFT) Plot
Yv = fft(vowel); Yc = fft(consonant);
f = (0:N-1)*(FsV/N);
magV = abs(Yv)/N; magC = abs(Yc)/N;

% === Frequency Spectrum (Separate, Improved Plot) ===
figure;
subplot(2,1,1);
plot(f(1:N), magV(1:N), 'b', 'LineWidth', 30.5);
title('Frequency Domain - Vowel (A)'); xlabel('Frequency (Hz)'); ylabel('Magnitude
|X(f)|'); grid on;

subplot(2,1,2);
plot(f(1:N), magC(1:N), 'r', 'LineWidth', 30.5);
```

```
➢ title('Frequency Domain - Consonant (H)'); xlabel('Frequency (Hz)');
  ylabel('Magnitude |X(f)|'); grid on;
➢
➢
➢ % Step 6: Spectral Centroid (Math)
➢ centroidV = sum(f(1:N/2)'.*magV(1:N/2)) / sum(magV(1:N/2));
➢ centroidC = sum(f(1:N/2)'.*magC(1:N/2)) / sum(magC(1:N/2));
➢
➢ fprintf('\nSpectral Centroid (Effort Measure):\n');
➢ fprintf('Vowel (A): %.2f Hz\n', centroidV);
➢ fprintf('Consonant (H): %.2f Hz\n', centroidC);
➢
➢ % === Equation Reminder ===
➢ % FFT: X(k) = sum x(n) * exp(-j*2*pi*k*n/N)
➢ % Spectral Centroid = sum(f * |X(f)|) / sum(|X(f)|)
```

- **<u>Audio Files Used:</u>**

File Name	**Type**	**Description**
A.m4a	Vowel	Pronunciation Of "A"
H.mp4a	Consonant	Pronunciation Of "H"