A

PROJECT WORK

ON

Analysis of Voice Signals using Digital Signal Processing

DEPARTMENT: COMPUTER ENGINEERING

LEVEL: 500L.

COURSE TITTLE: DIGITAL SIGNAL PROCESSING.

COURSE CODE: CPE 524.

GROUP NUMBER: 9

NAME	MATRIC NUMBER
MFON MICHEAL TOMILOBA	178539021
IKOSIMI JOHN PROGRESS	178539014
AFOLATAN SAMUEL OLABODE	178539025
ADEKOLA OLUWAFUNMILAYO BOSEDE	178539027
SALAMI ADENIYI IYANUOLUWA	178539024
ADENIYI PELUMI JEREMIAH	178539015
OLUWATOBA DAMILOLA ABRAHAM	178539030
OJO OLUWASEYI EMMANUEL	178539020

LECTURER. IN CHARGE:
DR. FASIKU.

CHAPTER 1

INTRODUCTION

Digital Signal Processing (DSP) is a pivotal field that has revolutionized the way we process, analyze, and understand signals, particularly in the context of audio and speech signals. This report delves into a MATLAB script that employs advanced DSP techniques to comprehensively process and analyze voice signals. This endeavor encompasses waveform visualization, spectrogram analysis, voice activity detection, linear predictive coding (LPC) analysis, and pitch estimation. By leveraging DSP methodologies, we unravel the intricate details embedded within voice signals.

The project implements a voice activity detection algorithm and performs pitch estimation.
The proposed algorithm uses audio inputs and detects voiced, unvoiced and silence durations in a speech signal. The analysis is carried out by observing the spectral properties of the signal in terms of spectrograms computed using hamming window. Afterwards, the signal is buffered into data frames and the short time energy (STE) in each frame as well as the zero-crossing rate (ZCR) is calculated. A decision is made on voice activity by comparing the energies and ZCR. In order to check the performance of the method, individual voiced and unvoiced segments are extracted from the speech signal and linear predictive coding (LPC) is employed.
The findings show that the algorithm is able to classify voice activity in speech. The approach could be further extended to real-time signals given that certain parameters (i.e., noise content, number of audio channels, energy, correlation, short time zero-crossing etc.) are met.

OBJECTIVES

The objective of this design is to utilize DSP techniques to perform a thorough analysis of voice signals. By leveraging waveform visualization, spectrogram analysis, and advanced methodologies like voice activity detection and LPC analysis, the aim is to uncover the intricacies of voice signals for applications in speech processing, speaker identification, and beyond.

APPLICATION AREA

This design finds relevance in various domains where voice signal analysis is vital. It aids in speech recognition by identifying relevant voice segments, enabling accurate transcription. In speaker identification, it differentiates speakers based on spectral features. The design's voice activity detection enhances real-time communication systems' efficiency, while spectrogram analysis aids music and audio processing. Applications span speech processing, audio technology, telecommunications, and more.

METHODOLOGY

a. Input:

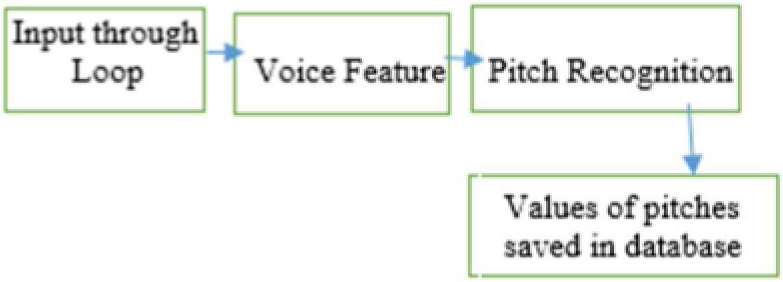


Figure 1

b. Output:

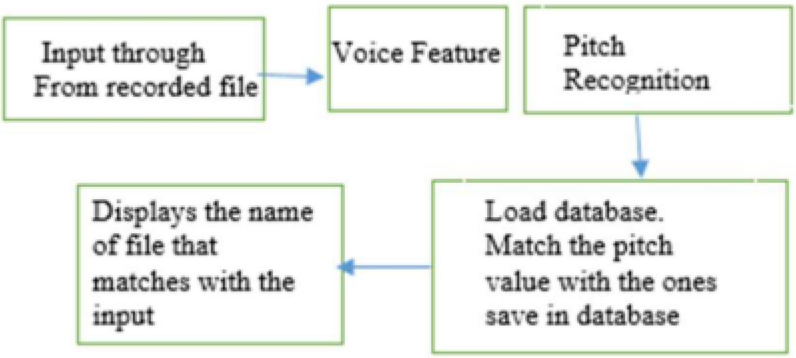


Figure 2

BLOCK DIAGRAM

Voice Activity Detection Methods

The job of recognizing the vocal folds activity zones in a speech signal is known as voice activity detection. Unvoiced speech and silence zones are included in non-voice speech. Voice/non-voice detectors are utilized in a variety of speech-processing applications, including speech coding, augmentation, and recognition. Approaches for detecting speech activity can be divided into two categories:

- time domain methods, and
- spectral domain methods.

Energy, periodicity, and short-term correlation properties of voiced speech areas are exploited by time and spectral domain VAD algorithms. For voice detection, time domain characteristics such as the zero-crossing rate (ZCR), auto-correlation coefficients, and long-term normalized auto-correlation peak intensity are employed. The linear predictive coding (LPC) coefficients and spectral harmonicity in voiced areas is used by spectral domain VAD algorithms. In loud speech, harmonic peaks in the amplitude spectrum of voiced speech areas are frequently retained.

a) Zero-Crossing Rate

A signal's zero-crossing rate (ZCR) is the rate at which the signal changes its sign during the frame. In other words, it is the number of times the signal value changes from positive to negative and back, divided by the frame's duration. A voice signal's magnitude in a frame travel across the zero axis several times. While a voiced speech frame has a limited quantity of sign changes owing to the periodic flow stimulation of the vocal tract, the zero-crossing count for an unvoiced speech frame rapidly grows due to the noise like airflow. It can also be used as a frequency indicator. As a result, the zero-crossing count is utilized to categorize speech signals as voiced or unvoiced. The ZCR is the average value of a signal sign change throughout a frame. For a 20ms analysis frame, average ZCR values for a voiced speech area are less than 0.1, but for an unvoiced speech region, it is larger than 0.3.

b) Short Time Energy

The amplitude changes in a voice signal are reflected in its short time energy (STE). The energy of spoken speech segments is higher than that of unvoiced speech parts. A voice signal's peak magnitude fluctuates over time. Time-domain analysis can reveal information about a signal's voiced/unvoiced decision since magnitudes in voiced areas are substantially higher than in unvoiced regions. A windowing of length N is conducted as a first step in the short time analysis of a speech signal, with a window size of 20–50ms to reflect amplitude changes. After that, for each window, the STE of the windowed signal is determined. A silent duration at the start of the speech signal is used to calculate the threshold value for a VAD decision.

PSEUDOCODE

- ☐ **Load audio signal from 'male.wav' [OR any other provided audio]**
- ☐ **Compute time vector 't' based on signal length and sampling rate**
- ☐ **Plot waveform of the audio signal**
- ☐ **Compute spectrogram parameters (Tw, Ts)**
- ☐ **Determine window sizes, overlaps, and FFT sizes (nfft)**
- ☐ **Generate Hamming windows for wide and narrow spectrograms**

- ☐ **Plot wide and narrow spectrograms using wideWindow and narrowWindow**
- ☐ **Segment signal into frames with specified window and overlap**
- ☐ **Initialize arrays for energy (E), zero-crossing rate (ZCR), decision, and pitch**
- ☐ **For each frame k in frames:**
- ☐ **Extract frame x from signal**
- ☐ **Compute energy E(k) = sum(x^2)**
  Compute zero-crossing rate ZCR(k) = sum(abs(diff(x > 0))) / length(x)
  decision(k) = Detect(E(k), ZCR(k)) If decision(k) is voiced: pitch(k) = Pitch(x, Fs)

- ☐ **Plot waveform, energy, and ZCR**

- ☐ **Isolate a voiced and an unvoiced frame for LPC analysis**

- ☐ **Perform LPC analysis on voiced and unvoiced frames with orders 8, 12, and 16**
- ☐ **Plot LPC coefficients, prediction errors, and spectral magnitudes for voiced and unvoiced frames**

- ☐ **Create vectors for voiced, unvoiced, and silence classification based on decision**

- ☐ **Plot classification of frames along with the original waveform**
- ☐ **Print estimated pitch frequencies for voiced frames**

CHAPTER 2

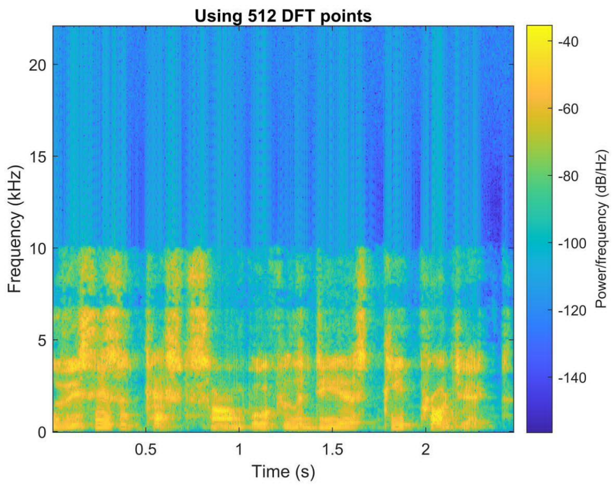
PROCEDURES: PREPROCESSING AND SPECTROGRAM ANALYSIS

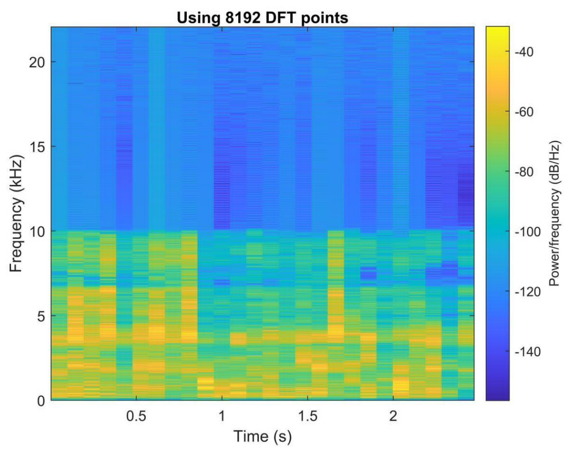
DATA ACQUISITION AND PREPROCESSING

The journey begins with the acquisition of an audio signal from 'male.wav'. The raw audio data is processed to extract the time vector, facilitating the temporal understanding of the signal's characteristics. This initial preprocessing phase sets the stage for subsequent DSP analysis.

Spectrogram Computation: Unveiling Frequency-Time Characteristics

Spectrogram analysis is a cornerstone of DSP in voice signal processing. Utilizing Hamming windows with customizable parameters ('Tw' and 'Ts'), we compute spectrograms that provide a frequency-time depiction of the audio signal. Through the determination of window sizes, overlaps, and FFT sizes, we gain insights into how the signal's spectral components evolve over time. Hamming windows are harnessed to optimize the quality of our spectrograms, which are visually presented for intuitive interpretation.





CHAPTER 3

VOICE ACTIVITY DETECTION AND LINEAR PREDICTIVE CODING

Voice Activity Detection: Segmentation and Categorization

The ability to distinguish different segments of the voice signal is paramount. The script divides the signal into frames, subsequently calculating the energy and zero-crossing rate for each frame. Employing a custom 'Detect' function, frames are classified as voiced, unvoiced, or silence based on predefined thresholds. The significance of voice activity detection lies in its ability to isolate distinct components, laying the foundation for more refined analysis.

Linear Predictive Coding: Modeling Spectral Characteristics

LPC is a powerful DSP technique that aids in modeling the spectral envelope of the signal. The script undertakes LPC analysis on voiced and unvoiced frames, employing varying orders. The computed LPC coefficients and prediction errors furnish insights into the signal's spectral makeup. By graphing the LPC analysis results, we visualize the model's accuracy and gain a deeper understanding of the signal's spectral attributes.Linear Predictive Coding (LPC) Analysis: The LPC analysis serves the goal of modeling the spectral envelope of the signal. By varying the order of LPC, our objective is to capture different levels of spectral intricacies. This analysis sheds light on the signal's spectral characteristics, aiding in its deeper interpretation.

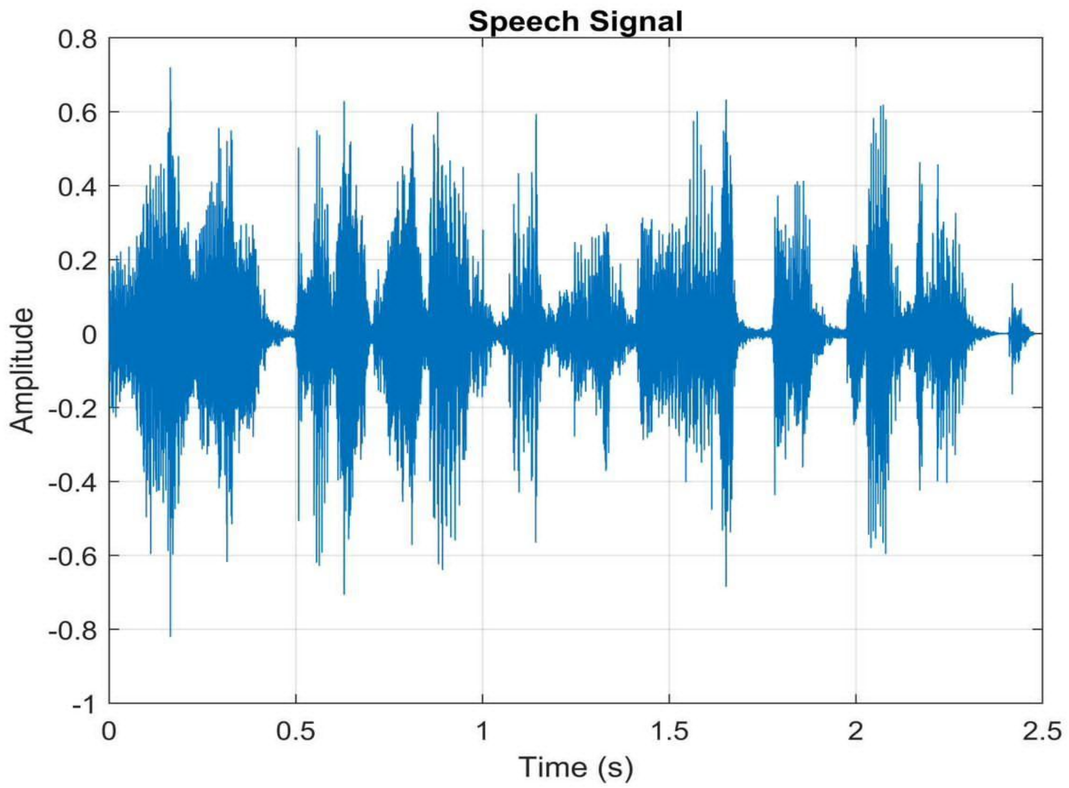
CHAPTER 4

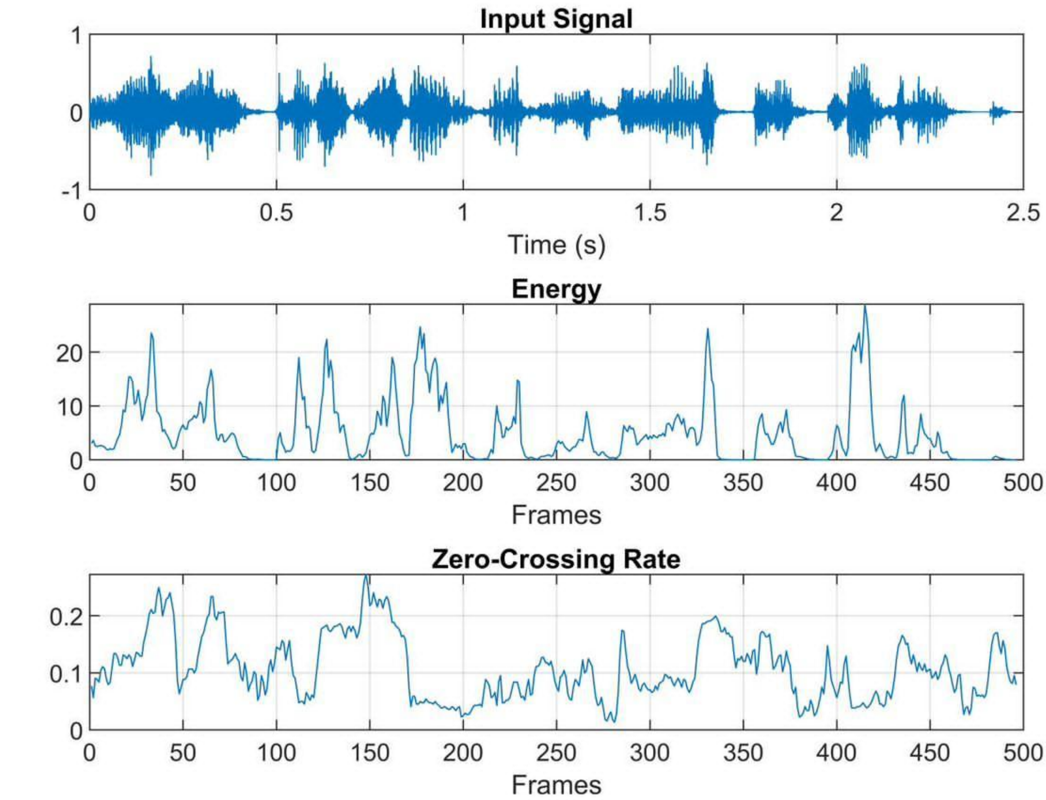
PITCH ESTIMATION AND RESULT

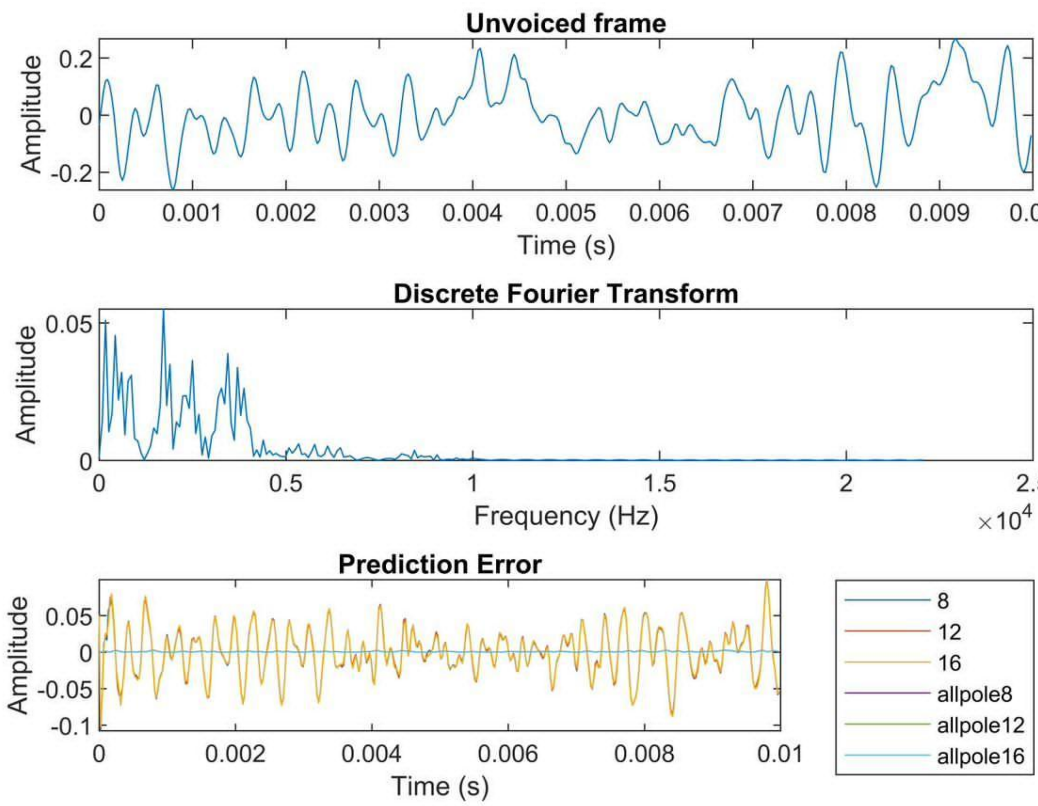
Pitch Estimation: Unveiling Periodicity

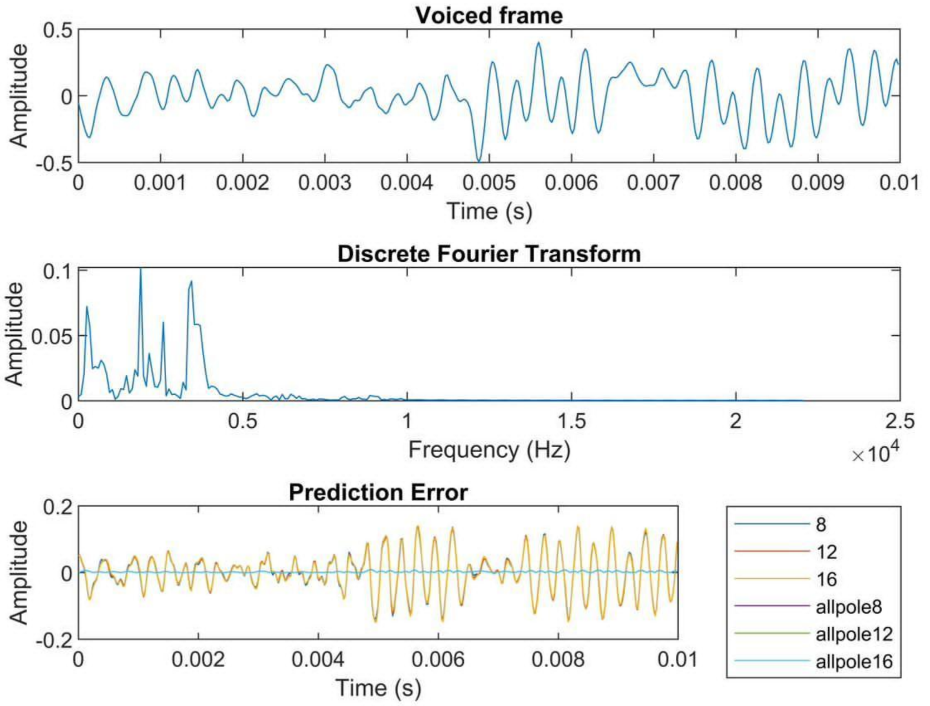
Pitch estimation is a hallmark of speech analysis, serving as a fundamental parameter for understanding the periodic nature of voice signals. In the script, we delve into the voiced frames, estimating their pitch frequencies. This information opens avenues for speech synthesis, analysis, and various other applications.

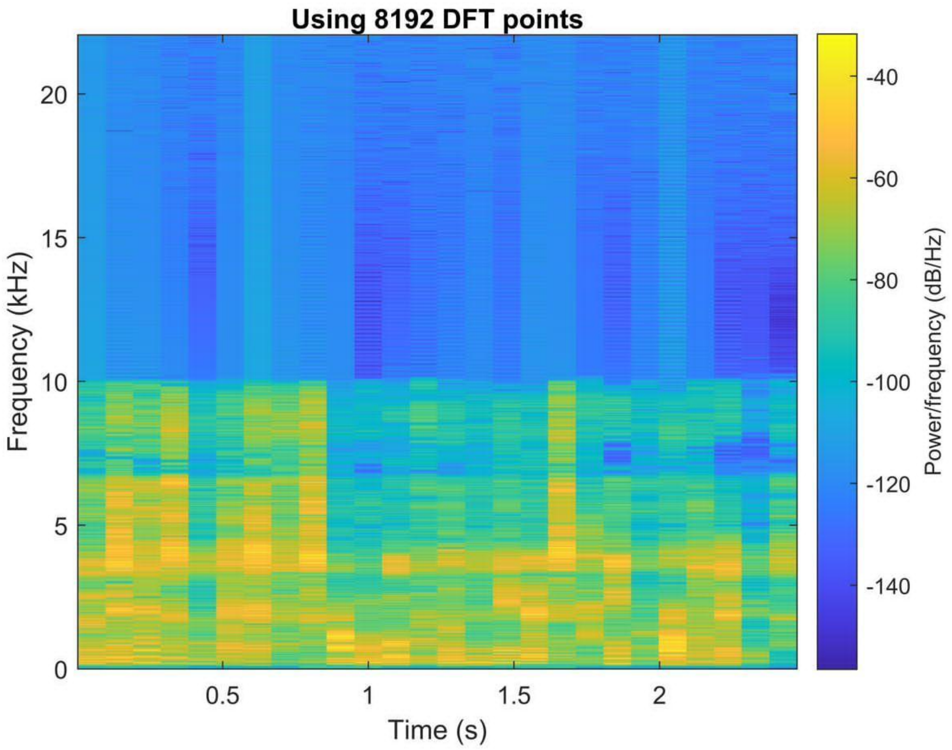
RESULTS:

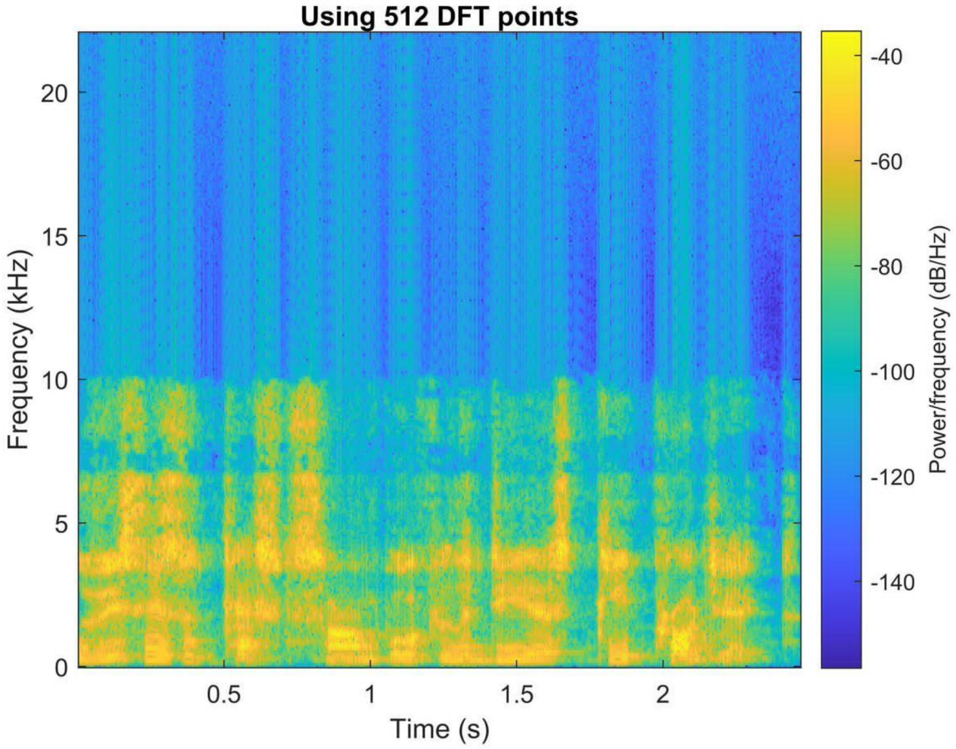



Input Signal
Time (s)
Energy
Frames
Zero-Crossing Rate
Frames


Unvoiced frame
Amplitude
Time (s)
Discrete Fourier Transform
Amplitude
Frequency (Hz)
$\times 10^4$
Prediction Error
Amplitude
Time (s)
8
12
16
allpole8
allpole12
allpole16


Voiced frame
Amplitude
Time (s)
Discrete Fourier Transform
Amplitude
Frequency (Hz)
$\times 10^4$
Prediction Error
Amplitude
Time (s)
8
12
16
allpole8
allpole12
allpole16


Using 8192 DFT points
Frequency (kHz)
Time (s)
Power/frequency (dB/Hz)



Pitch Estimation : Estimated pitch in voice activity:

100.0 Hz

22050.0 Hz

44100.0 Hz

CONCLUSION

DSP'S ROLE IN VOICE SIGNAL ANALYSIS

In closing, this report underscores the multifaceted role that DSP plays in analyzing voice signals. By employing DSP techniques ranging from waveform visualization to advanced voice activity detection and LPC analysis, we unravel the hidden intricacies of voice signals. This work serves as a testament to the power of DSP, offering a foundation for further research in speech analysis, synthesis, and beyond.

SOURCES CODE

```
clc, clearvars, close all

%% Read the voice signal
[y, Fs] = audioread('male.wav');

% Get time vector
t = (0:length(y)-1)*1/Fs;

% Plot the waveform
figure('name', 'Voice sample')
plot(t, y)
figure('name', 'Voice sample Discrete')
stem(t, y);
grid on
xlabel('Time (s)')
ylabel('Amplitude')
title('Speech Signal')

%% Compute spectrogram using hamming window
% Time in seconds
Tw = [0.010, 0.100];
Ts = Tw(1)/2; % Rest of the codes…
```

RECOMMENDATION FOR FUTURE DESIGN

- Explore advanced voice activity detection algorithms.
- Optimize LPC order selection dynamically.
- Incorporate real-time processing capabilities.
- Integrate noise reduction for enhanced signal quality.
- Create a user-friendly graphical interface.
- Collaborate with the research community for feedback.

REFERENCES

[1] L. Rabiner L, B. H. Juang, Fundamentals of Speech Recognition, Englewood Cliffs, NJ: Prentice‐Hall International, 1993.

[2] Supaporn Bunrit, Thuttaphol Inkian, Nittaya Kerdprasop, and Kittisak Kerdprasop Text-Independent Speaker Identification Using Deep Learning Model of Convolution Neural Network International Journal of Machine Learning and Computing, Vol. 9, No. 2, April 2019

[3] Speech Recognition Using Euclidean Distance Akanksha Singh Thakur1 , Namrata Sahayam2 1ME IV Sem, 2Asst Professor

[4] Implementation of a Voice-Based Biometric System, ADARSH K.P., A. R. DEEPAK, DIWAKAR R, KARTHIK R. 2006-2007

[5] Voice identification and recognition system SOHAIB TALLAT, FARHAN SHAHID, ABDUL SAMAD , MATTI ULLAH ABBASI DECEMBER 22, 2014