Mini-Project Presentation

PHONETIC TRANSCRIPTION WITH PROSODY ANALYSIS

BY

ARAVIND C (92132223014)

JAINUL ABUDEEN S (92132223058)

JEEVANNESWAR N (92132223061)

Abstract

- Phonetic transcription and prosody analysis are essential tools accurate representation of both speech & sound.
- Prosody analysis, on the other hand, captures elements such as intonation, stress, rhythm, and timing.
- The integration of prosody with phonetic transcription offers significant value in speech recognition technology.
- By offering, web-based platforms, create more accurate speech recognition systems.

Introduction

- Phonetic transcription provides a systematic way to visually represent the sounds of speech using IPA.
- Prosody—the rhythm, stress, intonation, and timing.
- Prosody adds layers of meaning, emotion, and structure to spoken language.
- Phonetic transcription fall short in representing these suprasegmental features comprehensively.
- This led to a growing prosodic elements into phonetic transcription to better capture the full range of spoken language characteristics.

Problem Statement

- It does not adequately reflect the prosodic features—such as intonation, stress, rhythm, and timing.
- This systems like IPA provide a methodfor representing the individual sounds of spoken language.
- The lack of detailed prosodic information in traditional phonetic transcription creates several challenges across various fields:
- Incomplete Linguistic Representation, Barriers in Language Learning:, Limitations in Speech Technology.

Challenges

- **Complexity of Representation:** Capturing features of speech requires a more elaborate transcription system,.
- **Subjectivity in Prosody Annotation:** Prosody, unlike phonetic sounds, is highly variable and context-dependent.
- **Technological Limitations:** Current speech recognition and transcription technologies are primarily focused on segmental sounds.
- **Lack of Standardization:** While phonetic transcription IPA, there is no universally accepted system for prosodic transcription.
- **Training and Usability:** prosodic features into phonetic transcription demands extensive training for linguists, language educators, and speech pathologists.

Tools & Technologies

Frontend:

- HTML , CSS for UI
- JavaScript for connectivity & interaction

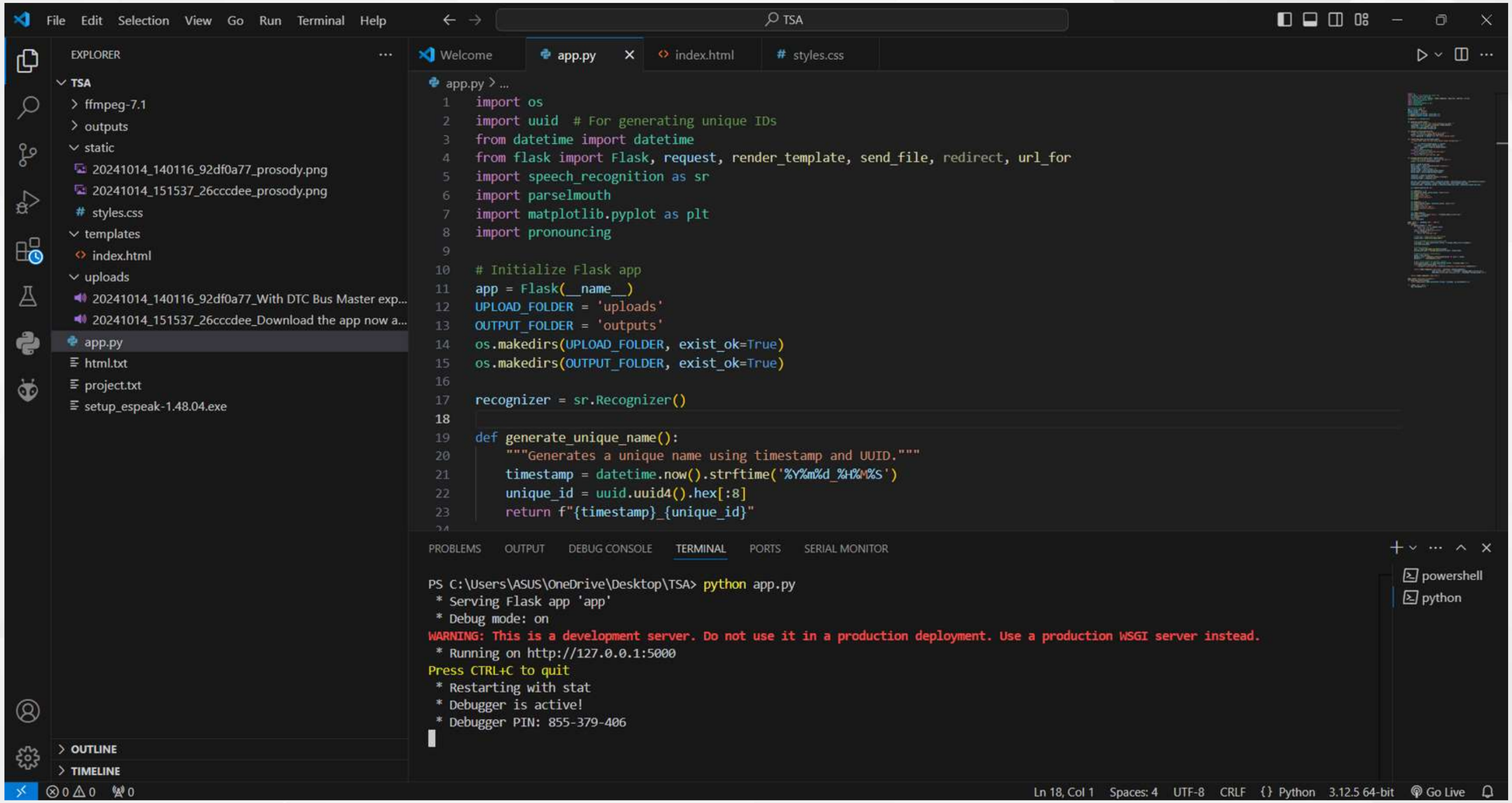
Backend:

- Python (Flask)
- Matplotlib for visualization

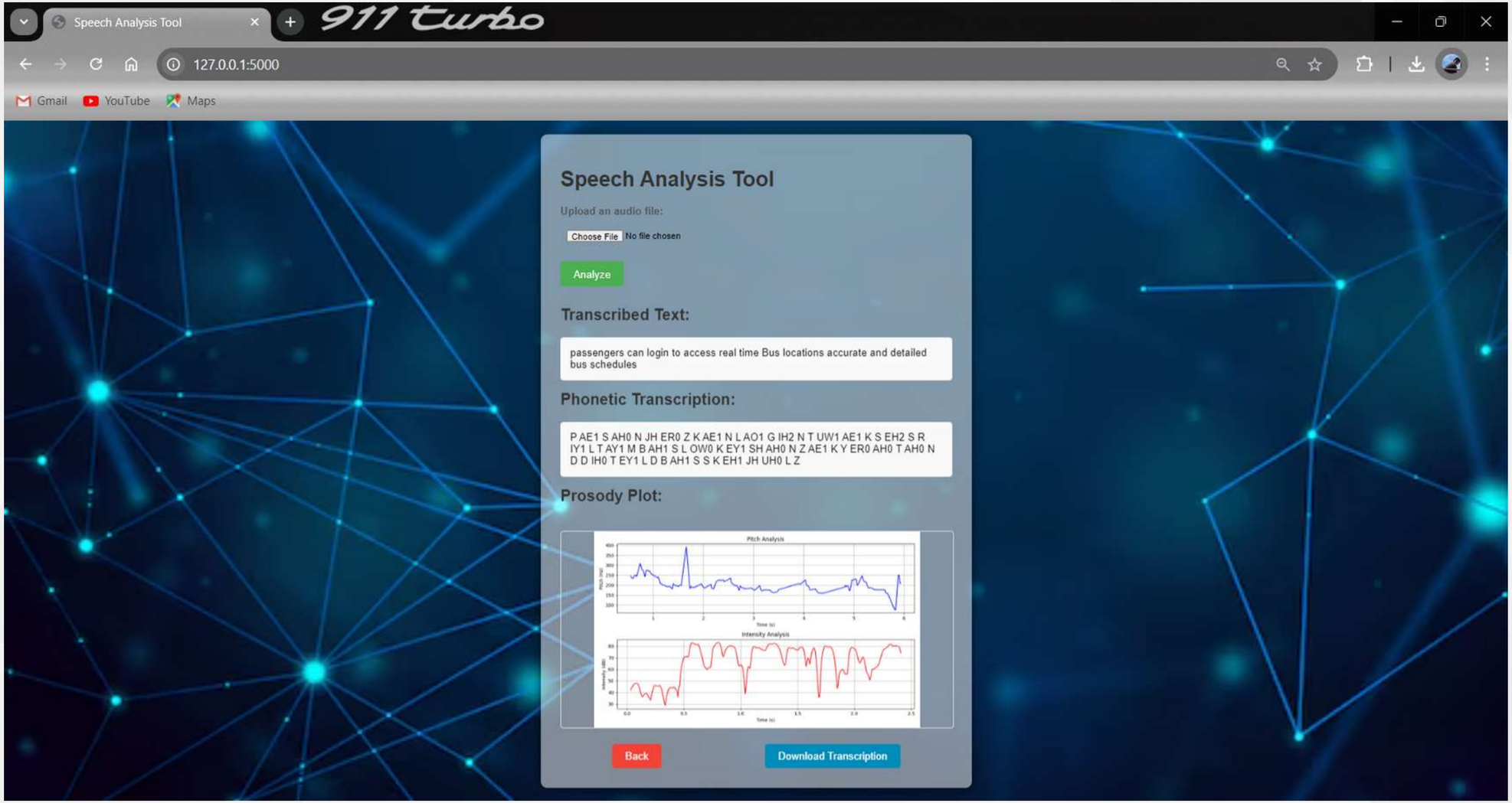
Proposed Model

- **Input**: Raw speech data (audio).
- **Phonetic Transcription Module**: Converts speech into phonetic symbols using machine learning models.
- **Prosody Analysis Module**: Extracts prosodic features like intonation, rhythm, and stress patterns using time-series analysis or deep learning techniques.
- **Phonetic Accuracy**: Evaluating how accurately the model transcribes speech into phonetic symbols compared to human experts or a standard dataset

Implementation



Demo Screenshort



Conclusion

- Phonetic transcription and prosody analysis are integral to understand spoken language.
- prosody analysis adds it’s output in the visualization.
- various applications, from linguistic research and language education to speech technology and clinical practice

PHONETIC TRANSCRIPTION WITH PROSODY ANALYSIS

A PROJECT REPORT

Submitted by

ARAVIND C	**92132223014**
JAINUL ABUDEEN S	**92132223058**
JEEVANNESWAR N	**92132223061**

MINI-PROJECT: PHONETIC TRANSCRIPTION WITH PROSODY ANALYSIS

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY

(An Autonomous Institution, Affiliated to Anna University, Chennai)

DINDIGUL - 624622

OCTOBER 2024

PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY,

(Autonomous Institution Affiliated to Anna University, Chennai)

DINDIGUL – 624622

BONAFIDE CERTIFICATE

Certified that this idea report "**PHONETIC TRANSCRIPTION WITH PROSODY ANALYSIS**" is the bonafide work of **"ARAVIND C (92132223014), JAINUL ABUDEEN S (92132223058), JEEVANNESWAR N (92132223061)**" who carried out the idea work under my supervision in filing the patent work.

SIGNATURE	SIGNATURE
Dr. A. VINCENT ANTONY KUMAR, M.E, Ph.D., **HEAD OF THE DEPARTMENT** PROFESSOR & HEAD DEPARTMENT OF IT PSNA COLLEGE OF ENGINEERING TECHNOLOGY, DINDIGUL -624622	**Dr. P. PRIYADHARSHINI M.E,** **SUPERVISOR** ASSISTANT PROFESSOR DEPARTMENT OF IT PSNA COLLEGE OF ENGINEERING TECHNOLOGY, DINDIGUL -624622

ABSTRACT

Phonetic transcription and prosody analysis are essential tools in the study of speech, facilitating accurate representation of both speech sound and the suprasegmental features that shape meaning and emotion in spoken language. Phonetic transcription systems like the International Phonetic Alphabet (IPA) offer a standardized method for documenting the precise articulation of speech across languages.

Prosody analysis, on the other hand, captures elements such as intonation, stress, rhythm, and timing, which are crucial for conveying pragmatic information and speaker intent. Integrating these prosodic features into phonetic transcription enhances our understanding of language patterns, enriching fields such as linguistics, speech technology, and language education.

This report explores the role of prosody in communication, reviews transcription systems that incorporate prosodic elements, and examines the challenges and benefits of combining phonetic and prosodic transcription. The integration of prosody with phonetic transcription offers significant value in speech recognition technology, clinical speech pathology, and language teaching by providing a more holistic approach to the analysis of spoken language.

By offering, web-based platforms can leverage combined phonetic and prosodic transcriptions to create more accurate speech recognition systems, advanced language learning tools, and improved diagnostic methods in speech therapy. The integration of prosody into digital transcription provides a holistic understanding of speech, improving communication tools and applications in various fields.

INTRODUCTION :

In the study and analysis of spoken language, both phonetic transcription and prosody play crucial roles. **Phonetic transcription** provides a systematic way to visually represent the sounds of speech using standardized symbols, such as those from the International Phonetic Alphabet (IPA). It captures the segmental aspects of speech, like vowels and consonants, and is widely used across fields such as linguistics, language education, speech technology, and clinical practice.

However, speech is not just about the individual sounds but also about **prosody**—the rhythm, stress, intonation, and timing that influence how speech is perceived. Prosody adds layers of meaning, emotion, and structure to spoken language, impacting how listeners interpret everything from simple sentences to complex dialogues.

Despite the importance of prosody in spoken communication, traditional phonetic transcription systems often fall short in representing these suprasegmental features comprehensively. This has led to a growing interest in integrating prosodic elements into phonetic transcription to better capture the full range of spoken language characteristics.

PROBLEM STATEMENT:

While phonetic transcription is effective in documenting the precise articulation of speech sounds, it does not adequately reflect the prosodic features—such as intonation, stress, rhythm, and timing—that are critical for conveying meaning, intent, and emotional nuance in natural language.

This systems like the International Phonetic Alphabet (IPA) provide a comprehensive method for representing the individual sounds of spoken language, they fall short in capturing the prosodic features—such as intonation, stress, rhythm, and timing ,that critical to the full meaning of speech.

Prosody is essential for conveying emotions, speaker intent, emphasis, and the structure of discourse. The lack of detailed prosodic information in traditional phonetic transcription creates several challenges across various fields.

CHALLENGES:

1. **Complexity of Representation:** Capturing features of speech requires a more elaborate transcription system, making it harder to learn, use, and interpret.
2. **Subjectivity in Prosody Annotation**: Prosody, unlike phonetic sounds, is highly variable and context-dependent. Intonation, stress, and rhythm can vary between speakers intent.
3. **Technological Limitations:** Current speech recognition and transcription technologies are primarily focused on segmental sounds
4. **Lack of Standardization:** While phonetic transcription has standardized systems like the IPA, there is no universally accepted system for prosodic transcription
5. **Training and Usability:** prosodic features into phonetic transcription demands extensive training for linguists, language educators, and speech pathologists

PROPOSED SYSTEM:

To address the challenges of integrating prosody into phonetic transcription, a Proposed System is developed that combines phonetic transcription with automated and manual prosody annotation.

This system aims to be comprehensive yet user-friendly, improving the accuracy of speech representation while being practical for linguists, educators, speech therapists, and speech technology developers.

The proposed system aims to bridge the gap between phonetic transcription and prosody analysis by combining automated detection with manual adjustments, making it adaptable and efficient for a range of applications. Whether in linguistic research, speech technology, language learning, or clinical settings, this integrated system offers a more complete and user-friendly approach to capturing the full range of spoken language features, ultimately enhancing understanding and communication across domains.

SOURCE CODE:

BACKEND CODE:

PYTHON:

app.py > ...

```python
import os
import uuid  # For generating unique IDs
from datetime import datetime
from flask import Flask, request, render_template, send_file, redirect, url_for
import speech_recognition as sr
import parselmouth
import matplotlib.pyplot as plt
import pronouncing

# Initialize Flask app
app = Flask(__name__)
UPLOAD_FOLDER = 'uploads'
OUTPUT_FOLDER = 'outputs'
os.makedirs(UPLOAD_FOLDER, exist_ok=True)
os.makedirs(OUTPUT_FOLDER, exist_ok=True)

recognizer = sr.Recognizer()

def generate_unique_name():
    """Generates a unique name using timestamp and UUID."""
    timestamp = datetime.now().strftime('%Y%m%d_%H%M%S')
    unique_id = uuid.uuid4().hex[:8]
    return f"{timestamp}_{unique_id}"

def phonetic_transcription(word):
    """Get the phonetic transcription for a word."""
    phones = pronouncing.phones_for_word(word)
    return phones[0] if phones else "No transcription found"

def transcribe_audio_to_text(audio_path):
    """Transcribes audio to text using Google Speech Recognition."""
    try:
        with sr.AudioFile(audio_path) as source:
            audio = recognizer.record(source)
        text = recognizer.recognize_google(audio)
        return text
    except sr.UnknownValueError:
```

0 ⚠ 0 0 Ln 13, Col 26 Spaces: 4 UTF-8 CRLF Python 3.12.5 64-bit Go Live

FRONTEND CODE :-

HTML:

templates > index.html

```html
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Speech Analysis Tool</title>
    <link rel="stylesheet" href="{{ url_for('static', filename='styles.css') }}">
</head>
<body>
    <div class="container">
        <h1>Speech Analysis Tool</h1>
        <form method="POST" enctype="multipart/form-data">
            <label for="audio_file" class="upload-label">Upload an audio file:</label>
            <input type="file" id="audio_file" name="audio_file" accept="audio/*" required><br>
            <button type="submit" class="analyze-button">Analyze</button>
        </form>

        {% if text %}
        <div class="result-section sparkle-effect">
            <h2>Transcribed Text:</h2>
            <p class="text-output">{{ text }}</p>

            <h2>Phonetic Transcription:</h2>
            <p class="text-output">{{ phonetic }}</p>

            <h2>Prosody Plot:</h2>
            <img src="{{ plot_url }}" alt="Prosody Plot" class="prosody-plot">

            <div class="buttons">
                <a href="/" class="back-button">Back</a>
                <a href="{{ download_url }}" class="download-button">Download Transcription</a>
            </div>
        </div>
        {% endif %}
    </div>
</body>
</html>
```

0 ⚠ 0 0 Ln 17, Col 1 Spaces: 4 UTF-8 CRLF Django HTML Go Live

CSS:

static > # styles.css > body

```css
/* General Styles */
body {
    font-family: 'Arial', sans-serif;
    background-image: url(https://img.freepik.com/free-vector/background-realistic-abstract-technology-particle_23-2148431735.jpg);
    background-size: cover;
    background-position: center;
    margin: 0;
    padding: 0;
    height: 100vh;  /* Use full viewport height */
}

/* Container Styling */
.container {
    background-color: rgba(255, 255, 255, 0.5);  /* Semi-transparent background */
    padding: 30px;
    border-radius: 10px;
    box-shadow: 0 0 15px rgba(0, 0, 0, 0.5);
    max-width: 600px;
    width: 90%;  /* Responsive width */
    margin: auto;  /* Center horizontally */
    position: relative;  /* Position relative for proper alignment */
    top: 50%;  /* Position the container in the middle of the viewport */
    transform: translateY(-50%);  /* Adjust for centering */
}

/* Heading Styles */
h1 {
    margin-bottom: 20px;
    font-size: 2em;
    color: #333;
}

/* File Input Styles */
.upload-label {
    display: block;
    margin-bottom: 10px;
    font-weight: bold;
```

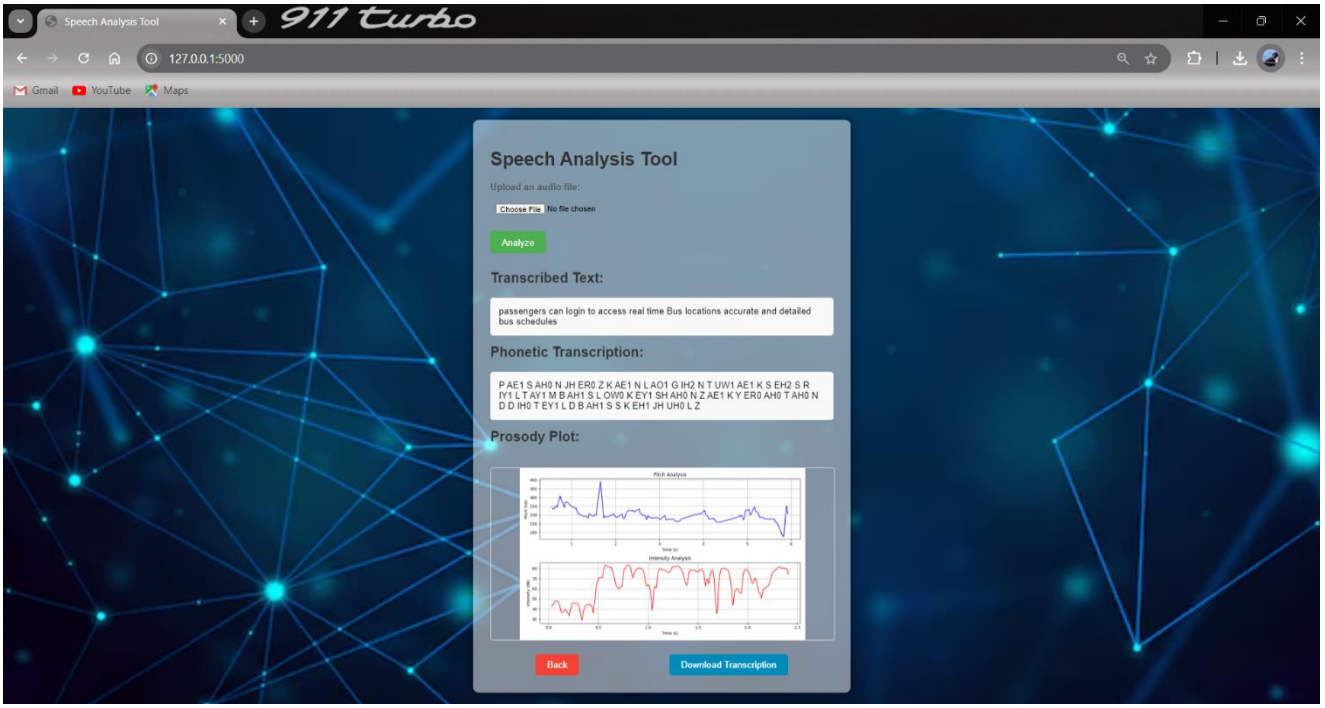
0 0 0 Ln 10, Col 2 Spaces: 4 UTF-8 CRLF CSS Go Live

static > # styles.css > body

```css
input[type="file"] {
    margin-bottom: 20px;
    width: 100%;
    padding: 10px;
}

/* Analyze Button */
.analyze-button {
    background-color: #4CAF50;
    color: white;
    padding: 10px 20px;
    border: none;
    border-radius: 5px;
    cursor: pointer;
    font-size: 1em;
}

.analyze-button:hover {
    background-color: #45a049;
}

/* Result Section */
.result-section {
    margin-top: 30px;
    text-align: left;
    animation: sparkle 1.5s ease-in-out forwards;
}

@keyframes sparkle {
    0% {
        opacity: 0;
        transform: scale(0.95);
    }
    50% {
        opacity: 1;
        transform: scale(1.05);
```

0 0 0 Ln 10, Col 2 Spaces: 4 UTF-8 CRLF CSS Go Live

OUTPUT:



CONCLUSION:

Phonetic transcription and prosody analysis are integral to understanding and representing the full spectrum of spoken language. While phonetic transcription captures the precise articulation of speech sounds, prosody analysis adds depth by revealing the rhythm, stress, and intonation that convey meaning and emotion. Integrating prosodic elements into phonetic transcriptions enhances various applications, from linguistic research and language education to speech technology and clinical practice