voisTUTOR - Speech Training Application

Ruthvik Kodati

2019101035

Aman Goyal

2019101097

Introduction

voisTUTOR is a mobile and web application which enables the user to improve their speech pronunciation by analysing their audio input. It delineates scores and graphs that are correlated to correctly uttered syllables and ones which were missed. Moreover, it also provides the expert audio output and the syllables within a phrase to help the user improve their pronunciation. We adhered to the following objectives:

1. Make a mobile application which interacts with the server and provides the user with an interface.
2. Making REST APIs that enable communication between the application and the server.

Motivation

Correct pronunciation is a basis for efficient communication in English as in any other language. Proper pronunciation can be defined as a reproduction of language sounds so that the intended message is efficiently passed and understood by a fluent speaker. Poor pronunciation can result in a failure to communicate your message and as a result, issues in comprehension. When an opportunity presents itself, you do not want poor English pronunciation to make a poor impression and ruin your chance for success.

We attempted to make an interface which is easy to use and learn. The interface is vastly interactive and correctly analyses the users' audio input. Ultimately, it displays relevant insights and results which would help the user improve their pronunciation.

Methodology

The aim was to create a mobile application which takes user input and lets the user decide between categories that give lessons to aid in overall speech improvement. Also, we had to make a server which uses REST APIs for communication to get user-friendly results as fast as possible.

Server:

Objectives Completed -

1. Understand the earlier server files which interacted using web sockets.
2. Configure the functions and methods to make them compatible with REST API.
3. Make various REST API endpoints to interact with the mobile application.
4. Process user queries and audio files and provide analysis results.

Tech Stack Used -

1. We used NodeJS and Javascript to make the server.
2. We used Express to make REST API.

Mobile Application:

Objectives Completed -

1. Make a Login and Register interface.
2. Display four categories for the user to select from.
3. Display different lessons present in each category for the user to select from.
4. Display various questions for each lesson. After the user input, they would receive the results.
5. All the screens were made using the screenshots of the previous mobile application.

Tech Stack Used -

1. We used React Native and Javascript to build the whole mobile interface.

Results

Backend API endpoints

Host : https://asr.iiit.ac.in/chiranjeevi/voisserve

GET

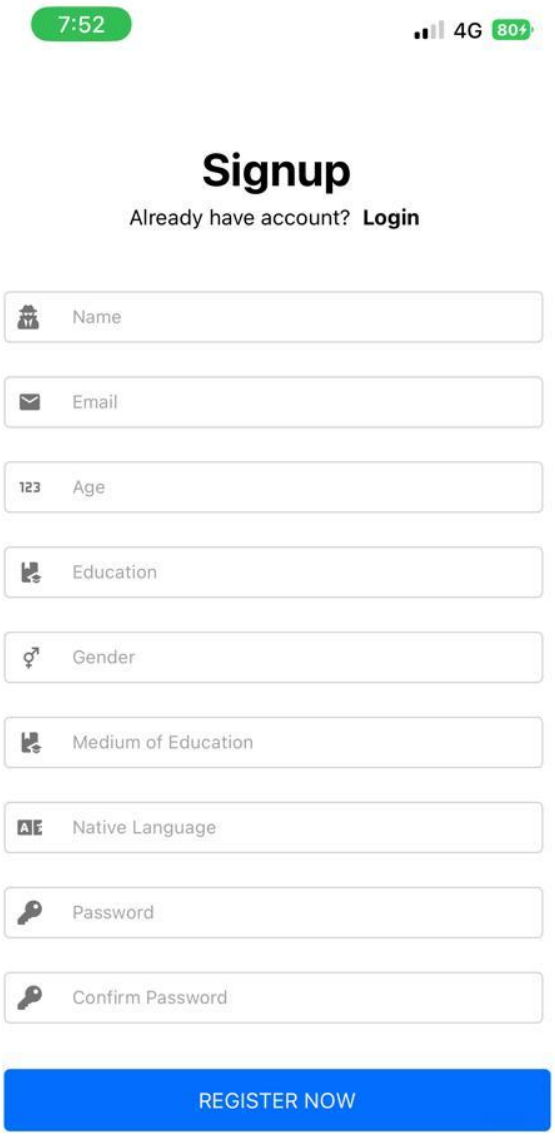
1. /get/categories : Get all the categories (Phoneme,Stress,Intonation,Sentence)

POST

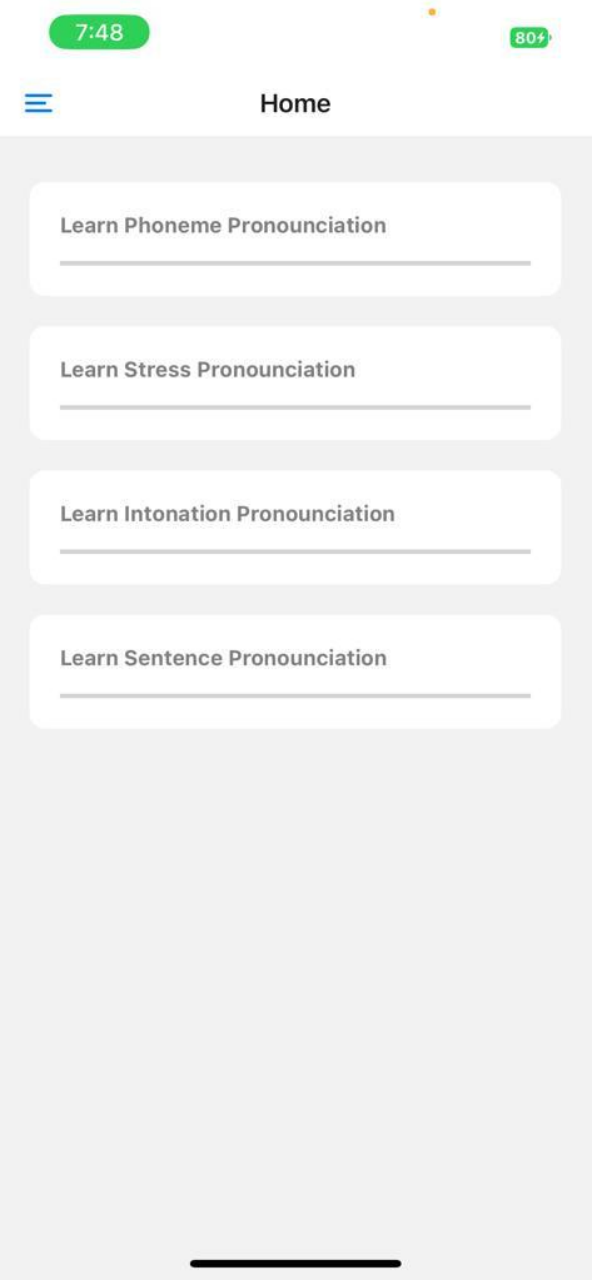
1. /post/login : Login user using email and password
2. /post/register : Register a new user using the required details
3. /post/updateLatestWord : Update the latest word chosen by the user
4. /post/lessons : Get all the lessons for a particular category using a category id
5. /post/questions : Get all the questions for a particular lesson
6. /post/phoneme: Post the audio files recorded for the phoneme category for analysis
7. /post/stress : Post the audio files recorded for the stress category for analysis
8. /post/intonation : Post the audio files recorded for the intonation category for analysis
9. /post/sentence : Post the audio files recorded for the sentence category for analysis

Frontend App

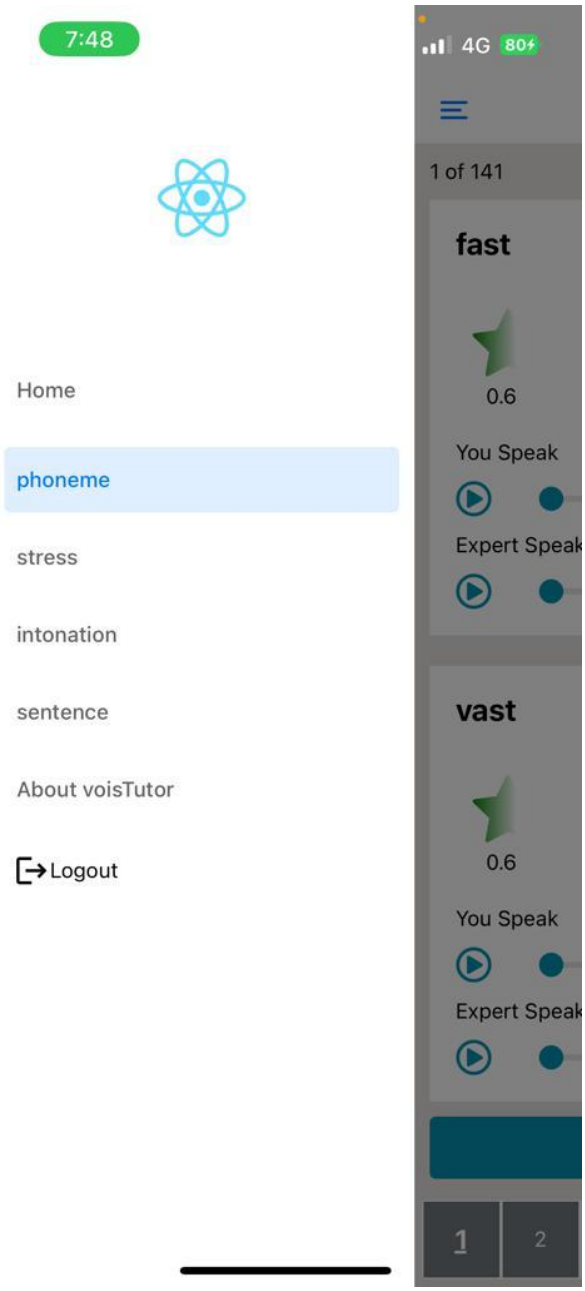
Login screen which takes in the email and password to login user



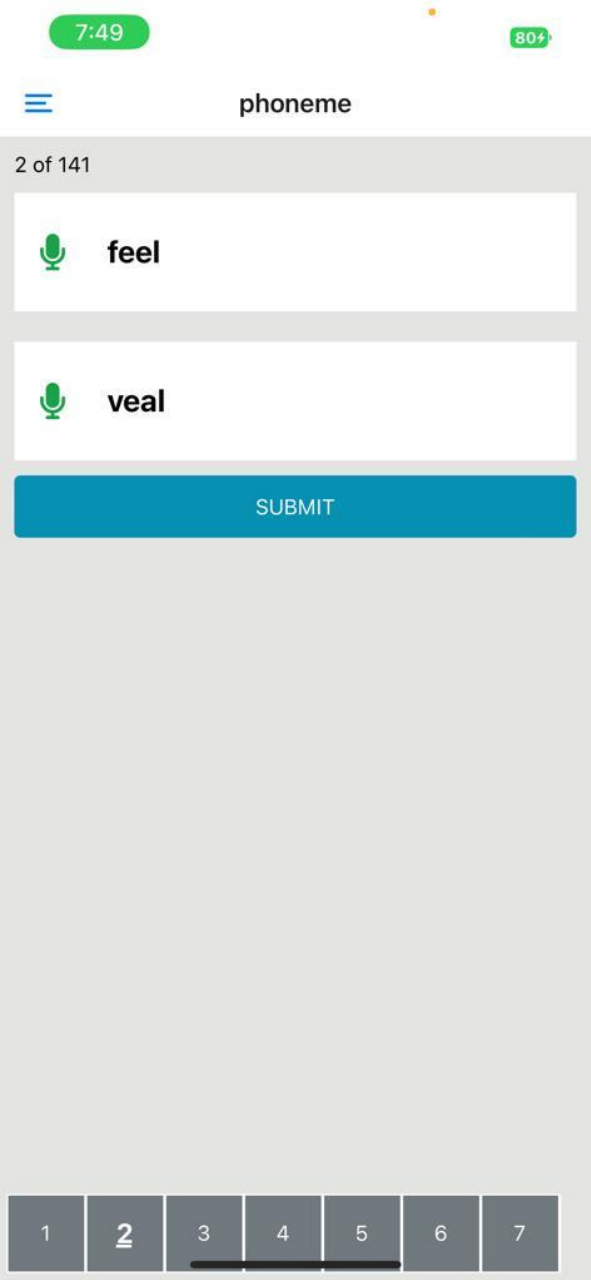
Signup screen which takes in the required details and registers a new user



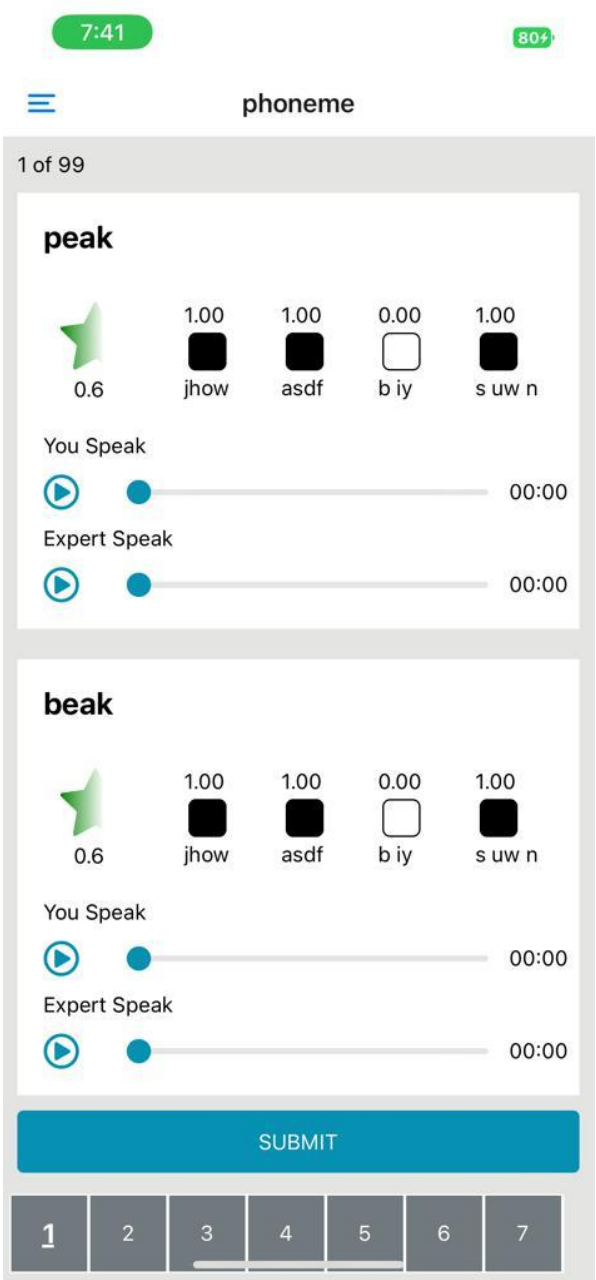
Category screen which lets the user select a category



Sidebar which eases navigation betweeen different categories and also lets the user logout



The above screen is a common for all the categories which lets the user record the audio for each question of thatl lesson



The above screen is a results screen showing the analysis results for the audio the user recorded and also plays the expert audio

Conclusion

Both of us worked on the project as part of our credit requirements for the B.Tech Project. We were interested in the problem statement and the learning we would get by making the project.

Our goals were to produce a robust server which processes user queries efficiently and to make an interactive mobile application which is clear-cut. We are content with the final deliverables and are hopeful that the project idea would prove to be fruitful for people seeking to ameliorate their English pronunciation.