UE20CS390A – Capstone Project Phase – 1

Project Progress ESA Review

Project Title : Voice Interface for PESU using AI
Project ID : PW23_PB_01
Project Guide : Prof. Priya Badrinath
Project Team with SRN : 805_806_826_844

Abstract and Scope

Abstract:

The voice assistant will be designed to help student with various academic and administrative task such as, finding courses, checking grades, accessing other resources like time table.

Scope:

The voice assistant will be developed using natural langue processing (NLP) and machine learning algorithm, which will enable it to understand and respond to user requests.

1. Easier access to PESU resources
2. Get the latest information about academics
3. Tap talk and get wherever you wish in PESU academy.

Suggestions from Review - 2

- Suggestions and remarks given by the panel members.

  - Fetching from documents: We are going to use OpenCV for extraction of table information from ISA and any PDF into CSV.

  - Data and Database: We are building our own database by using MySQL and the data we are getting from Kaggle dataset.

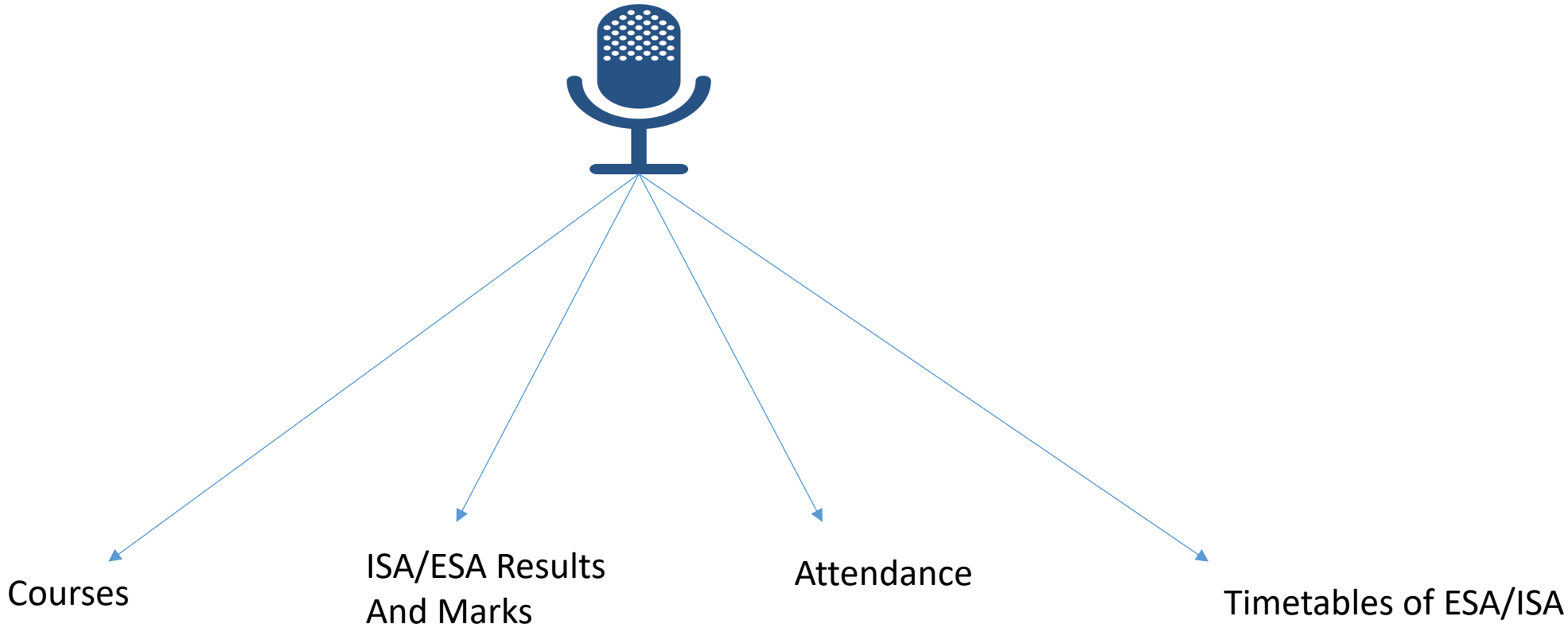
  - Conversion of voice to text: We will use python speech recognition library to convert voice to text

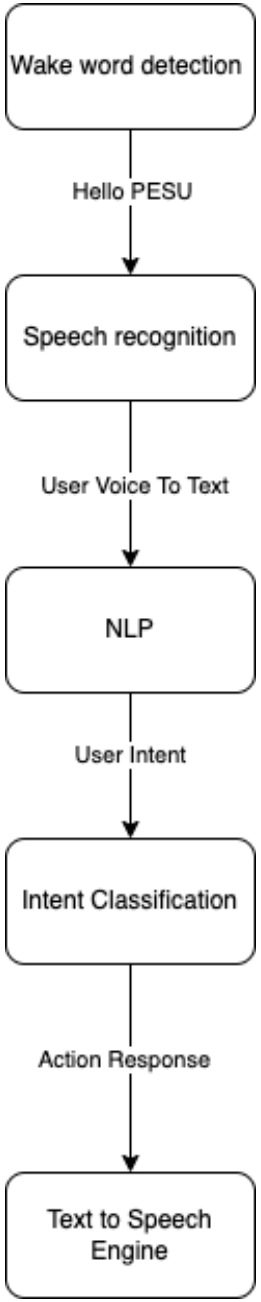
Suggestions from Review - 3

- Suggestions and remarks given by the panel members.

  - Panel members Raised a thought on complexity of the Project: - On this suggestion, We have decided to cut down some of features Like Voice Authentication and Web scraping.

Our key features now are



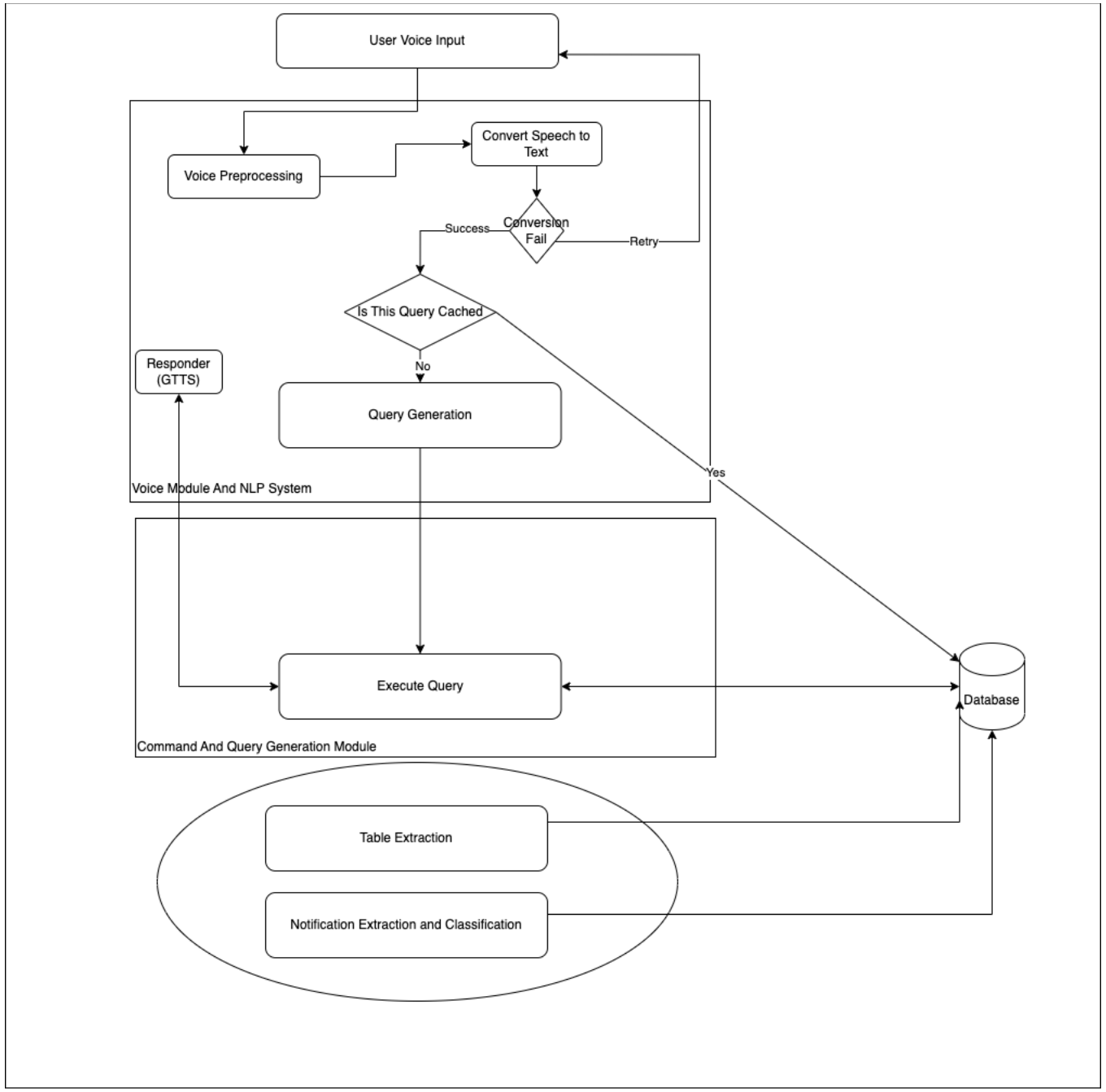


Design Approach

- Here We have opted Simple voice assistant Design Approach As Suggested by panel members consists of several key components
  - Including wake word detection
  - Speech recognition
  - Natural language processing (NLP)
  - Intent classification
  - Action generation
  - Text-to-speech.
- The assistant listens for a specific wake word i.e. “Hello PESU”, then records the user's voice and converts it into text using a speech recognition engine. The text is processed using NLP to understand the user's intent, and the intent is classified to determine the appropriate action or response. Finally, the response is generated and played back to the user in the form of audio using a text-to-speech engine.

NLP Module

- We use the Python SpeechRecognition library to convert the user's voice input into a desirable text format.
- Natural Language Processing (NLP) will be used to convert normal language queries into database queries by leveraging various techniques such as :
    - **Tokenization**
    - **Part-of-speech tagging**
    - **Named entity recognition**
    - **Query generation**
    - **Query execution**



Architecture

- Student Voice is given as input and it is sent to server
- Server pre-processes the voice for reducing noise, converts into intents and checks if it is cached or not
- If not, request is made to Command Module which takes intents and arguments as input and Generates Database Query, which is then executed and result is sent to user and narrator narrates the results
- For every time period, Table Extraction, Fetching of marks, Attendance are done

Literature Survey

- As You can see in image You need to upload documents.
- Trained model will Check for its quality. After converting It from pdf to image
- Then document scanner Will scan the document pre-process the image

Literature Survey

Data Extraction:

- In order to, extract data from image Image has to be in good quality. That's why we are using Logistic Regression to classify image as good or bad. Accuracy of Logistic Regression model was 83%

- For Extraction of data from. We are going to use OpenCV and PY tesseract for text extraction from each cell.

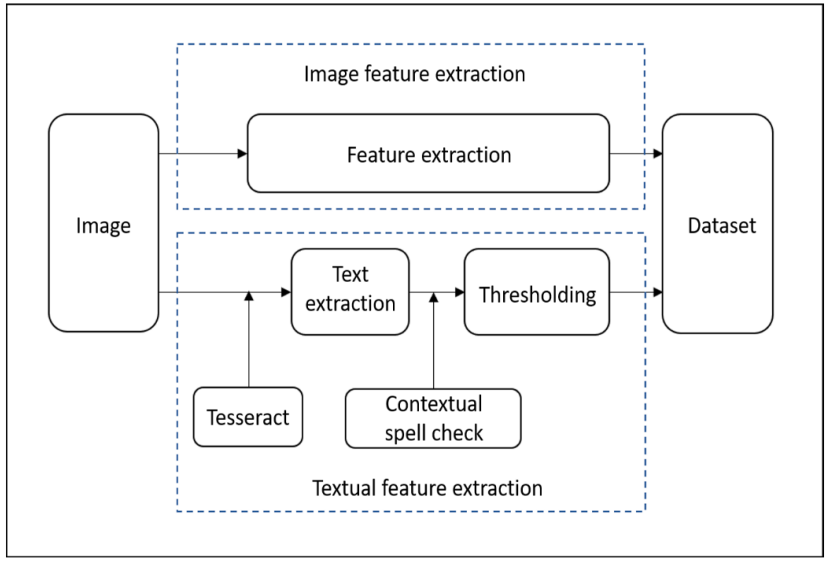


Fig. 3: Schematic flow of dataset compilation.

Literature Survey

- Then Cells are detected and text are extracted and csv/json/txt/excel is generated
- Logistic regression is used for quality checks
- Performance Is optimized by creating background mask is created by calculating the size of height and width from all the cell ROI crops
- OpenCV is also been used to perform Wrap perspective and rotation if needed

Table	Basic Approach	Background mask approach
Table with 24 cells	24 seconds	3 seconds
Table with 50 cells	50 seconds	4 seconds

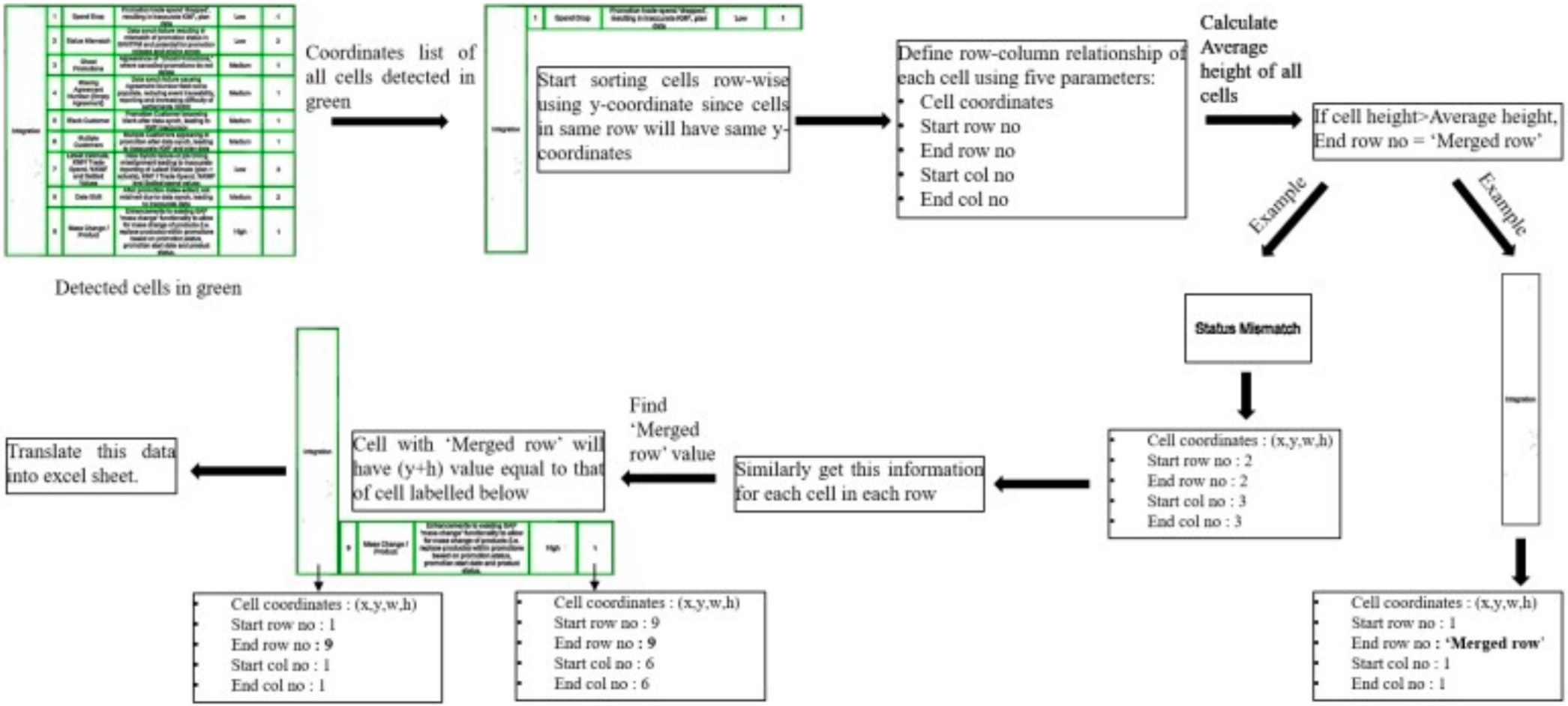
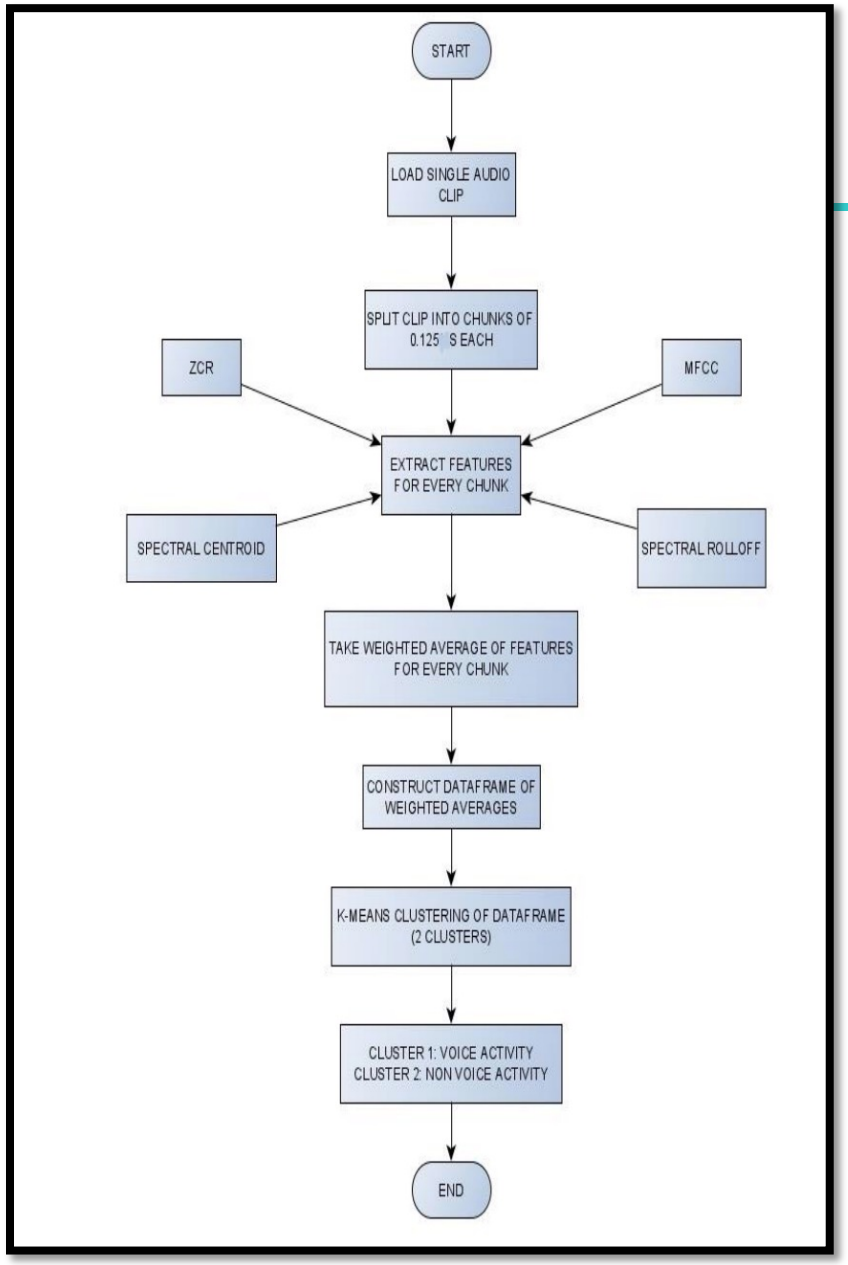


Fig. 6: (a) Cell extraction using x and y coordinates as explained above.

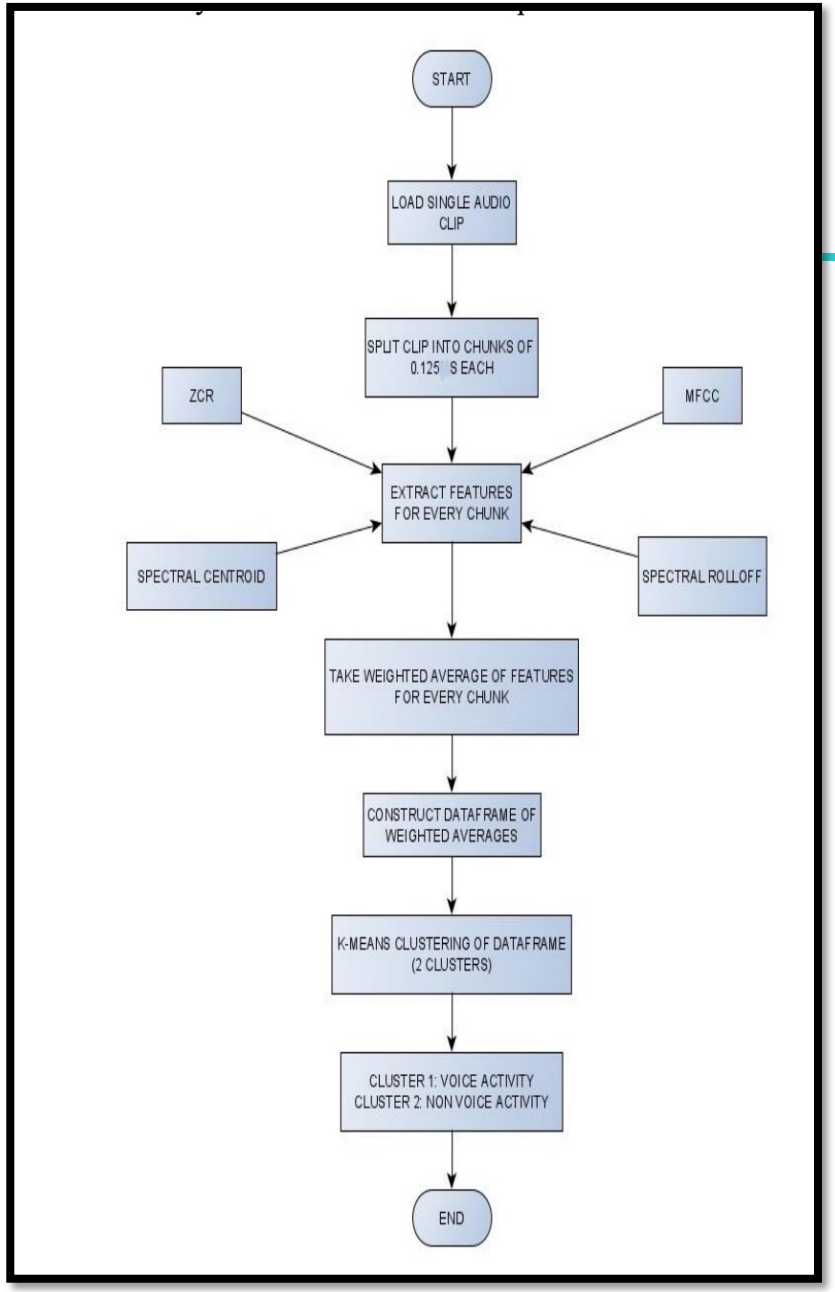
Literature Survey



- Working model is simple, audio is broken down into .125 seconds clip, fed to feature extraction algorithm
- Then, weights are added and passed to KNN, KNN classifies as human voice and non human voice

Audio Features : -

- Mel-Frequency Cepstral Coefficients
- Spectral Roll – Off
- Spectral Centroid
- Zero Crossing Rate



Literature Survey

What do these features tell us

1. The pitch of the sound: Spectral Centroid and MFCCs can help identify the pitch or fundamental frequency of the sound.

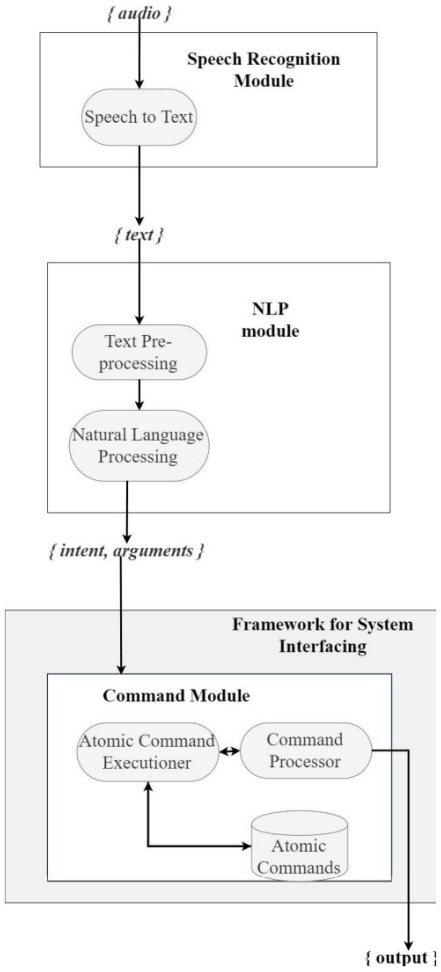
2. The timbre of the sound: MFCCs can help to identify the timbre or quality of the sound, such as whether it is a voice, a musical instrument or noise.

3. The rhythmic structure of the sound: Zero Crossing Rate can help to identify the rhythmic structure of the sound, such as how often it repeats.

4. The overall shape and texture of the sound: Spectral Roll-Off and Spectral Centroid can help identify the overall shape and texture of the sound, such as whether it has a lot of high-pitched or low-pitched frequencies.

Literature Survey

A Framework for System Interfacing of Voice User Interface for Personal Computers



- This paper aims at generating intents and arguments
- As shown in the image, text is given as input to the STT module
- NLP module generates intents and arguments
- Then pass to command Module, where the actions/operations are performed

Literature Survey

- The SpeechRecognizer library can be used in 2 variants.
- Online and offline modes, only difference is in online mode audio is passed to Certain Api’s
- On the other hand, offline mode makes use of the popular ’CMU Sphinx’ library for conversion of speech to text.

Literature Survey

Artificial intelligence based Voice assistant

- This paper goes in depth about how to get audio from user and process it with Acoustic Analysis
- Acoustic Modelling - it represents that the elements were pronounced or not and what are the words which can complete these elements.
- Pronunciation Modelling: That analyses the way, where how these elements are pronounced, it will check whether there is any accent or other peculiarities.
- Language Modelling: This is often aimed toward finding contextual probabilities counting on what elements were captured.

Summary of Literature Survey

- Models specified in these papers are having Accuracy rate of more than 94%

- According to the survey, We may be able to increase performance, due to the fact that there will be less noise compare to what they're using

- And for the pdf to csv conversion we may have to implement our own algorithm for enhanced preprocessing, because our pdf are slightly inclined due to little stutter, while capturing the document using phone camera

Project Progress

What is the project progress so far?

- Based on suggestions of panel members, we have decided to go with simple approaches.

What is the percentage completion of the project?

- Roughly 60 %

Upcoming phase

- The main thing of part of our project is NLP, We are going to build a model for converting student query into executable query.
- We are going to design a program to preprocess all the data that we have extracted from table.
- Next part of our project is going to be database design and security, that is encryption
- We have to look into handling of cached data.

References

- Artificial_Intelligence-based_Voice_Assistant - Subhash S, Ullas A, Santhosh B, Siddesh S
- A Framework for System Interfacing of Voice User Interface for Personal Computers Preet Dabre, Rohit Gonsalves, Raj Chandvaniya, and Anant V. Nimkar Department of Computer Engineering Sardar Patel Institute of Technology Mumbai, India
- Table Detection and Extraction using OpenCV and Novel Optimization Methods – Nidhi, Karandeep Saluja, Asmita Mahajan, Akash Jadhav, Nakul Aggarwal
- Web Scraping Approaches and their Performance on Modern Websites - Ajay Sudhir Bale, Naveen Ghorpade, Rohith S, S Kamales
- A Method for Voice Activity Detection using K-Means Clustering - Atul Rohit Agarwal, Sourabh Tiwari, Sudhakar M S, Sankar Ganesh S

Thank You