INDIAN INSTITUTE OF TECHNOLOGY, GUWAHATI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Project Report On

SPEECH BASED VLC CONTROLLER

Based on Speech Recognition System

Submitted To

Prof. P. K. Das

Submitted By

HARSHIL SADHARAKIYA (224101022)

UTSAV BHEDA (224101054)

VATSAL VASOYA (224101056)

For course fulfilment of CS566: Speech Processing

ACKNOWLEDGEMENT

This project is being submitted as a requirement for course fulfilment of CS566 - Speech Processing. It is a pleasure to acknowledge our sense of gratitude to **Prof. P.K. Das** who guided us throughout the project work. His timely guidance and suggestions were encouraging. We thank the Teaching Assistants who were always helpful in clearing doubts. Finally, we thank our classmates for their support.

1. Harshil Sadharakiya (224101022)
2. Utsav Bheda (224101054)
3. Vatsal Vasoya (224101056)

Table Of Content

1 Abstract

This project implements a speech based VLC controller using speech recognition techniques and it is developed using C language. Using this system, a user can control various types of features of VLC using speech and with no intervention of the keyboard. User just has to provide a dataset of his/her voice speaking those various commands in order to develop the speech based recognition system. The system uses Hidden Markov Model to identify the features of the speech which can be used to recognize speech when we deploy the system in the real world.

2 Introduction

2.1 Purpose

The purpose behind this project is to make the VLC media player handsfree where a user can control various features that VLC media player provides without touching the mouse or the keyboard. Using our system, a user can speak various commands to be executed in the VLC and the system recognizes these commands using speech recognition techniques and executes these commands in the VLC media player for the user. So a user can operate the VLC media player without using a mouse or a keyboard and only his/her voice.

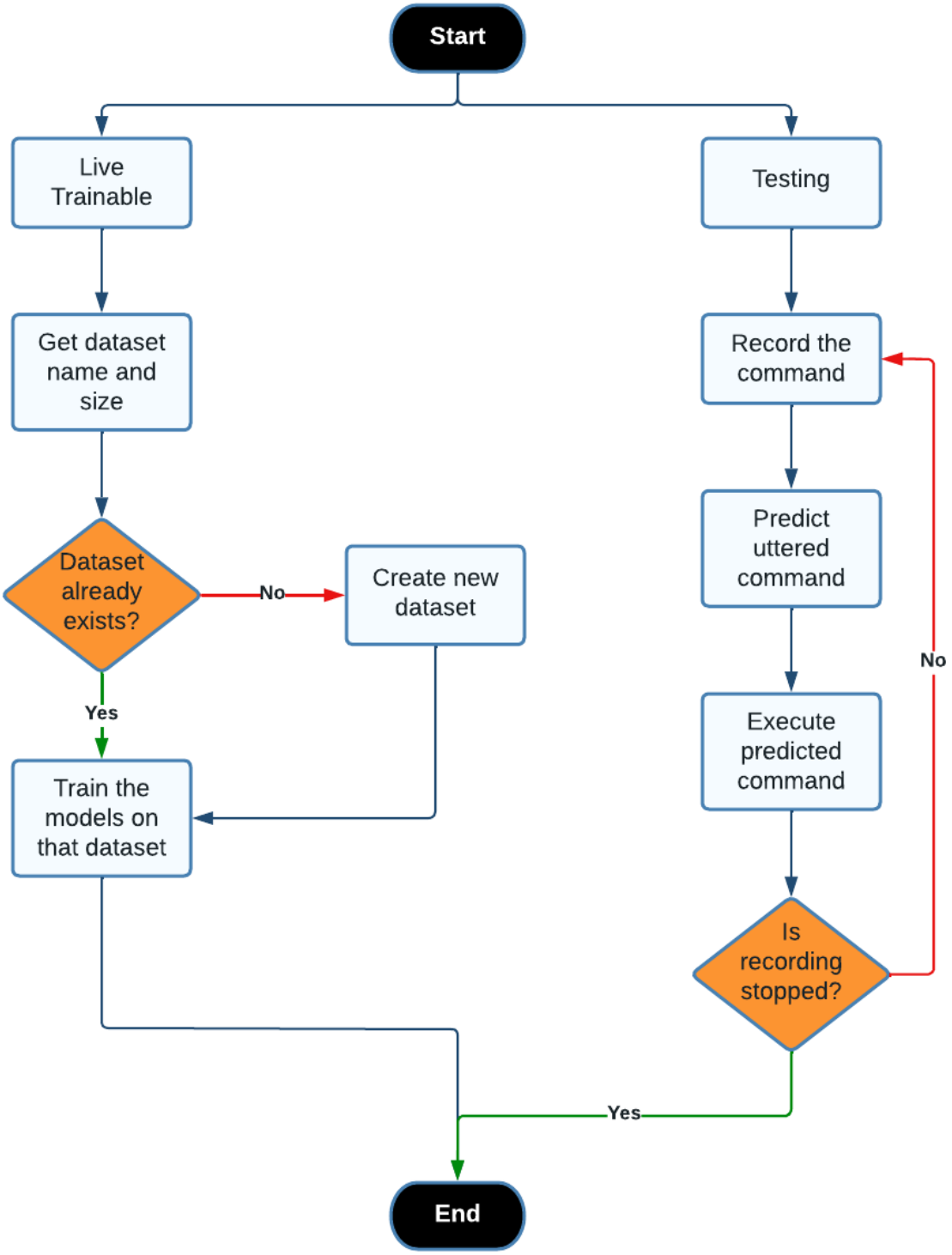
2.2 Our Project

This project builds a speech recognition system for a specific user. So in order to achieve that a user has to provide a dataset containing his/her voice speaking the specific words using which a command to handle a VLC feature can be generated. Then our system trains the Hidden Markov Models for all these words spoken by that specific user and builds a speech recognition system for that specific user. After the models are trained the model contains the meaningful feature information for the speech which can be used to recognize these specific words when it is spoken by that specific user.

2.3 Features

- List of words in the system:
  - Open
  - Movies
  - VLC
  - Volume
  - Window
  - Up
  - Down
  - Close
- List of commands supported by the system:
  - Open Movies
    - To open a folder named “Movies” which contains a list of videos which can be played by the user.
  - Window Open
    - To open the selected video from a folder containing a list of videos.
  - Window Close
    - To close the current active window.
  - Window Up
    - To move up on the list of videos inside the Movies folder.
  - WIndow Down
    - To move down on the list of videos inside the Movies folder.
  - VLC Close
    - To close the VLC media player.
  - VLC Volume Up
    - To increase volume in the VLC media player.
  - VLC Volume Down
    - To decrease volume in the VLC media player.

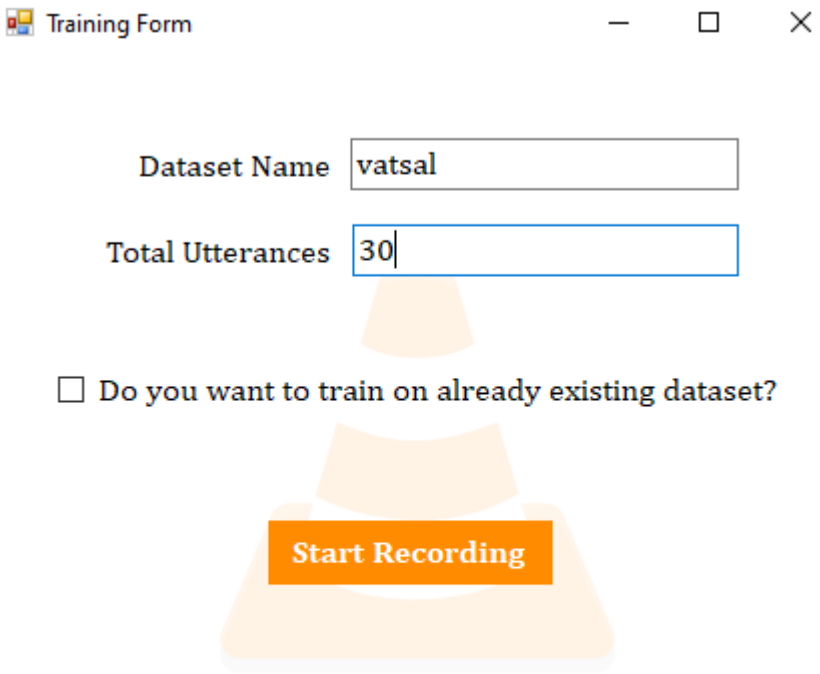
3 Flowchart


Start
Live Trainable
Testing
Get dataset name and size
Record the command
Dataset already exists?
No
Create new dataset
Predict uttered command
Yes
Execute predicted command
No
Train the models on that dataset
Is recording stopped?
Yes
End

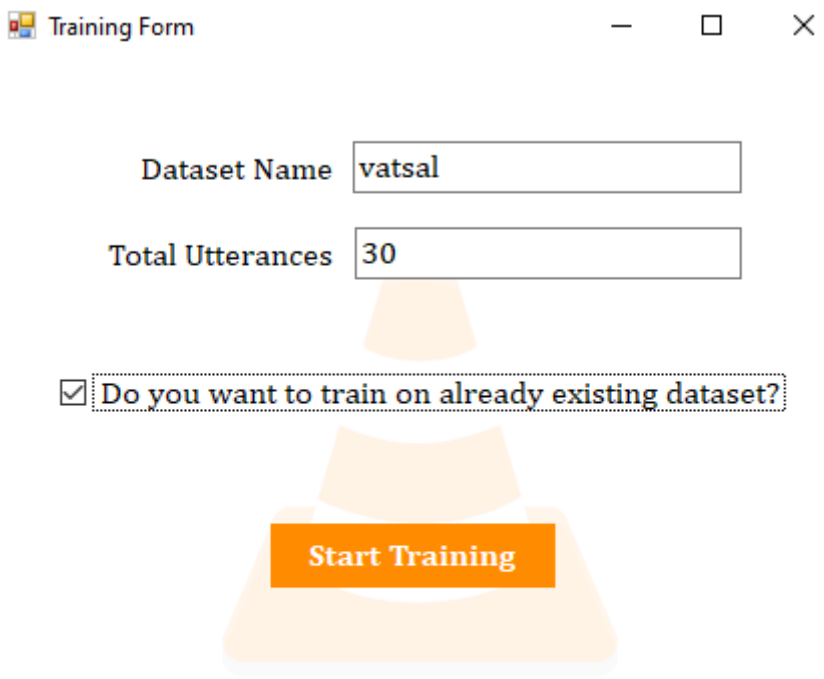
4 User Interface

4.1 Home Page

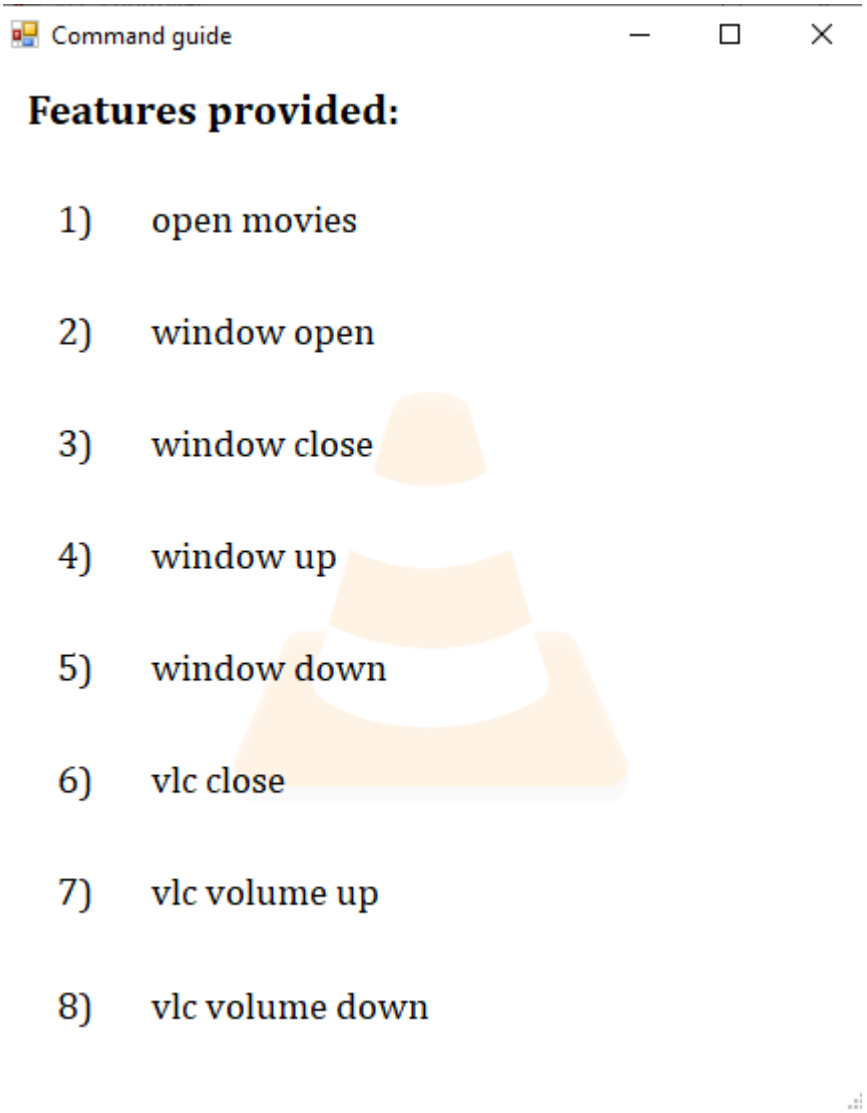
4.2 Trainable Module(When dataset is to be created)



4.3 Trainable Module(When dataset already exists)



4.4 Command Guide



5 Future Improvements

We can add more commands to use more features provided by the VLC media player using speech. To do that we can increase the number of words in the dictionary. For now our system provides speech based recognition for only a specific user and It is not that accurate if you use it for some other user so to improve our system, we can add data for various users to get the variety of speech data and learn the features so that our system can be used for any user rather than a specific user.