

# Indian

Institute

# of Technology,

Guwahati

# Department of Computer Science And Engineering

Project Report On

# Speech

Based

# PPT Manager

Based on Speech Recognition System

### Submitted To:

Prof. (Dr.) Pradip K. Das

### Submitted By:

Madhurima Sen (224101034)

Soumya Asati (224101046) Vaishali Chaudhari (224101055)

For course fulfillment of CS 566: Speech Processing

# Acknowledgement

This project is the submission for the course fulfillment requirement of the CS 566 - Speech Processing subject. We feel a deep sense of pleasure to acknowledge our gratitude to Prof. Pradip K. Das for providing us with an exciting project work experience in his subject. His timely guidance at every step throughout the project was encouraging and pushed our limits in completing our project. We would also like to thank the Teaching Assistants of the subject, who were always ready to clear our doubts. Finally, We are indebted to our classmates. Without their support, project completion within a strict deadline was not possible. As our unity motto, Apes Together Strong.

Madhurima Sen (224101034)

Soumya Asati (224101046)

Vaishali Chaudhari (224101055)

## Contents

1. [ABSTRACT](#_bookmark0) 4
2. [INTRODUCTON](#_bookmark1) 5
   1. [What is Speech Recognition?](#_bookmark2) 5
   2. [Our Project](#_bookmark3) 5
   3. [Future Improvements](#_bookmark4) 5
3. [EXPERIMENTAL SETUP](#_bookmark5) 6
4. [PROPOSED TECHNIQUES](#_bookmark6) 7
   1. [Flowchart](#_bookmark7) 7
   2. [Model description](#_bookmark8) 7
   3. [Modules](#_bookmark9) 7
5. [SNAPSHOTS](#_bookmark11) 10

## ABSTRAT

In this application, user can open, manage, close presentation saying some commands (hands free). The Application is developed in C/C++ in Visual Studio IDE.

List of commands used

1.Open : to open a specified presentation

2.Close : to close a specified presentation and it will also break out of the application

3.Back : to go to previous slide

4.next:to go to the next slide

5.goto:to go to a specific slide

6.present:to go to full screen

## INTRODUCTON

### What is Speech Recognition?

Speech recognition is an interdisciplinary sub-field of computer science and computational linguistics that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers. Some speech recognition systems require ”training” where an individual speaker reads text or isolated vocabulary into the system. Modern general-purpose speech recognition systems are based on Hidden Markov Models.

### Our Project

This project uses Hidden Markov Model to detect the spoken word out of the vocabulary present. Vocabulary in this Speech-based Project consists of some numbers (one, two, three, four, five, six and seven) and operations (Open, Close, Next, Present, Goto, Back). More operations can be further added to a fixed number, but the number of operations is currently fixed to six to show the developed prototype. Speaking the operation will detect that word and perform that particular operation on PowerPoint Application. The project includes facilities to train these words. It also includes live training of the words.

### Future Improvements

Future Improvements: The project is currently developed using C/C++ and is a bit system dependent. Future improvements include developing it into an independent system component, including more Applications, which are currently fixed as Microsoft PowerPoint. Also, the UI of the application can be made more dynamic and rich. Other multithreading supported High-level languages can be explored to run the training and testing components parallelly.

The project currently is unable to detect the sound continuously requiring the user to speak repeatedly and continuously instead of detecting when the speaker speaks the words for the operations on the PowerPoint presentation. This can be improved by exploring the data structures such as circular queues and detecting the words at real time.

## EXPERIMENTAL SETUP

Basic requirements for this project are as follows-

* Windows OS 10.
* Microsoft Visual Studio 2010.
* C++11 integrated with VS2010.
* Command Line Recording Module.
* A good Microphone

With the availability of above soft-wares, we further proceed in modelling the logic. The prerequisites of this project are:

* Basic i/o operations on File.
* Pre-processing of speech data. Generating Coefficients.
* Feature extraction.
* Modelling of extracted feature.
* Enhancing model.

## PROPOSED TECHNIQUES

### Flowchart

Figure [1](#_bookmark10) is a flowchart of the project. Flowchart can be referred for successful execution of the project.

### Model description

We are using the famous Hidden Markov Model for speech recognition. Hidden Markov Model is a probabilistic model used to derive the probabilistic characteristic of any random process. We use Cepstral Coefficients to represent the speech properties. We take all such cepstral coefficients generated by preprocessing the speech frames and build a codebook that helps in generating the observation sequences. The codebook contains 25 speech samples for each word.

We start with a feed-forward model and use the word observation sequences one by one to converge the model to its optimal value. Later on, we average out all the converged models of that word that save it to a hard disk.

While testing, we score each model using the Forward Process and pick the word with the highest score as a resultant word. Stress is present since speech signals depend significantly on the environment; therefore, live testing might not be excellent. However, we might get significantly better accuracy if we train the model live and test it immediately.

### Modules

Appropriate log files are generated for each operation in their respective folder which can be further referred for debugging purpose.

* + 1. **Observations Sequence Module:** It Generate Observation Sequence for each word and their all utterances for every training and testing files present.
    2. **HMM & Convergence Module:** It Converge All the Word Models one by one using the training observation sequences generated earlier. For the new word it generates observation sequences from it’s training files and then converge the model for that particular word only.
    3. **Testing Module:** It take the observation sequences of testing files and detect which word they belong to. It contains live testing of the word.

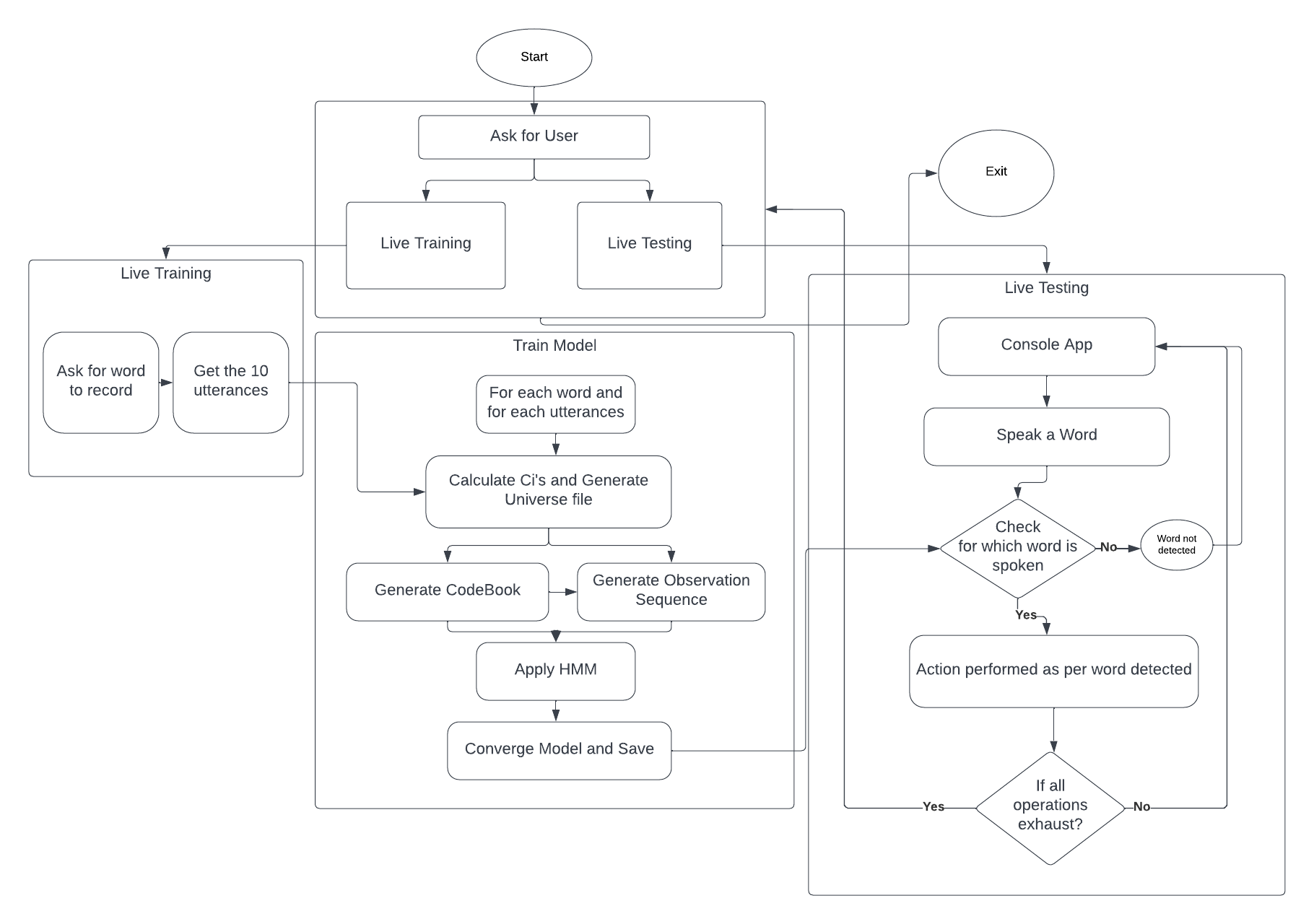


Figure 1: Flowchart of the Application

* + - * Online Live Testing involves speaking one utterance using recording module and then checking it which word system is detecting.
    1. **Training Module:** Here we can record the new utterances for the speaker which can be used for converging the module.
       - Live Training involves speaking one utterance of particular word then converging it by taking feed forward model. Later on averaging it with existing model of that word.
       - Recording Utterance Menu: where we can offline record each utterance of the all the words one by one. Later on it can be used for initial Convergence of the model.

## SNAPSHOTS

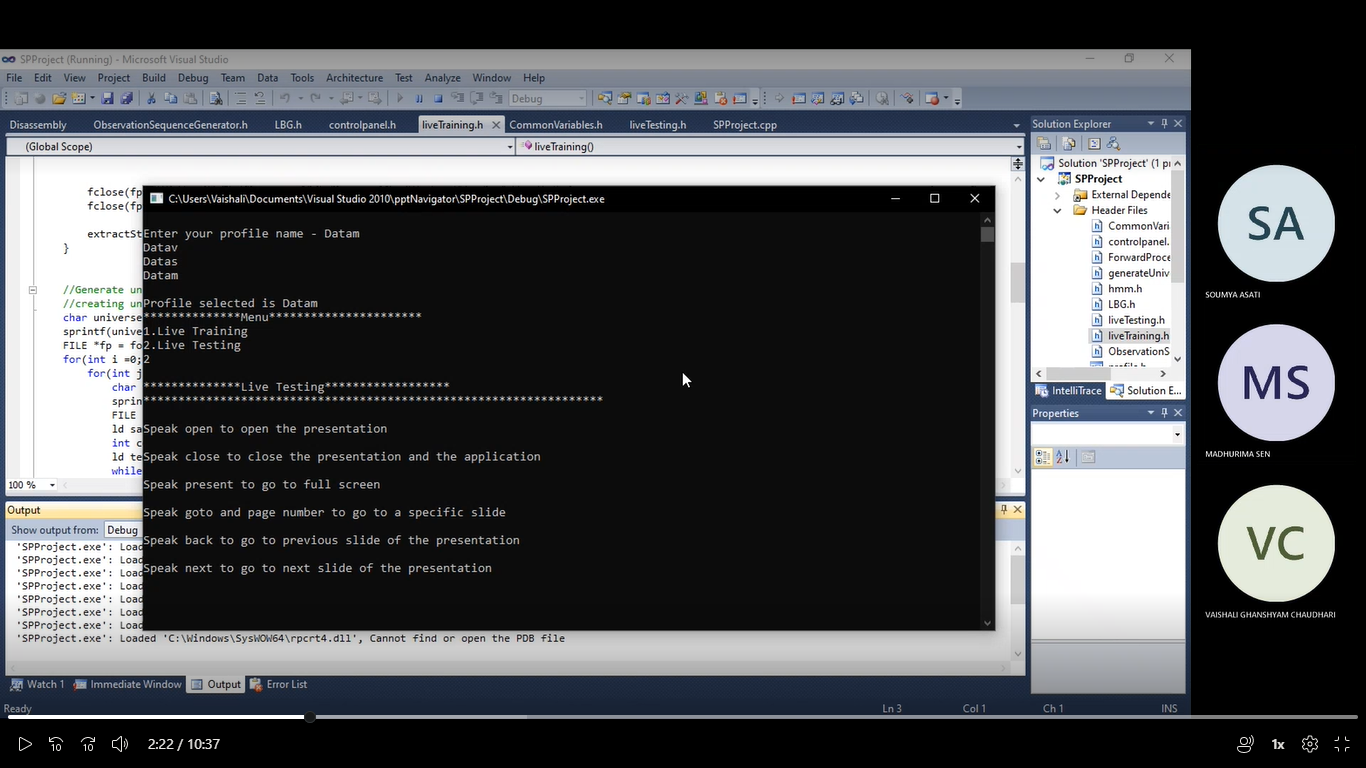


Fig. Main menu of the Application

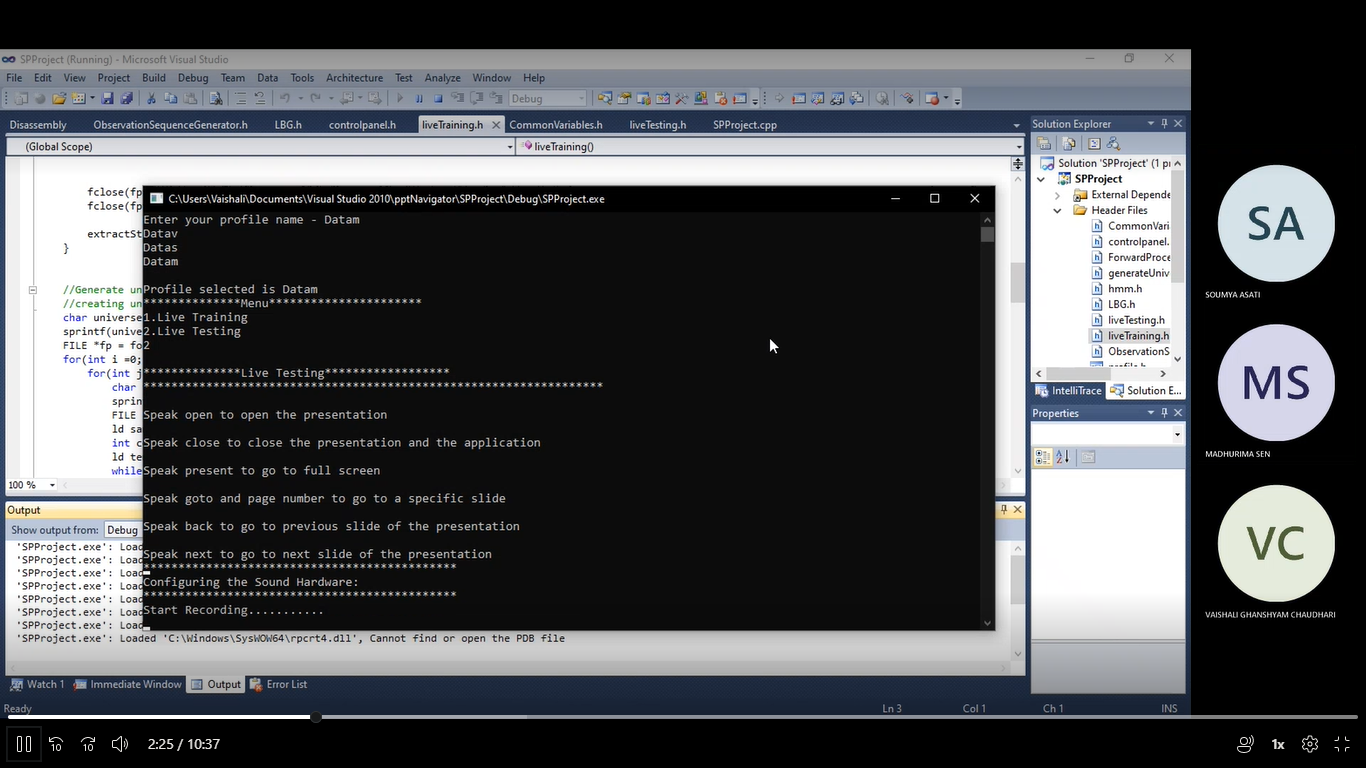


Fig. Recording live recording

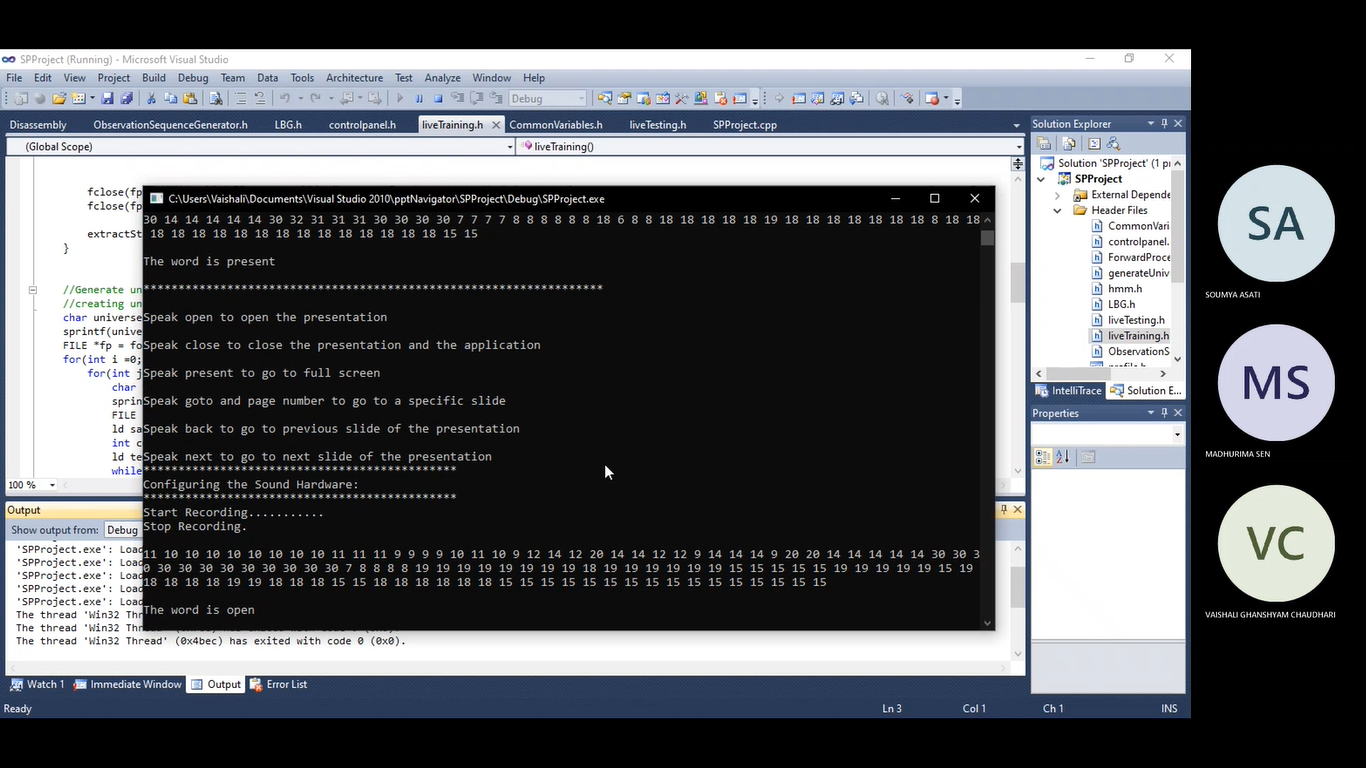


Fig. Open word is recognized

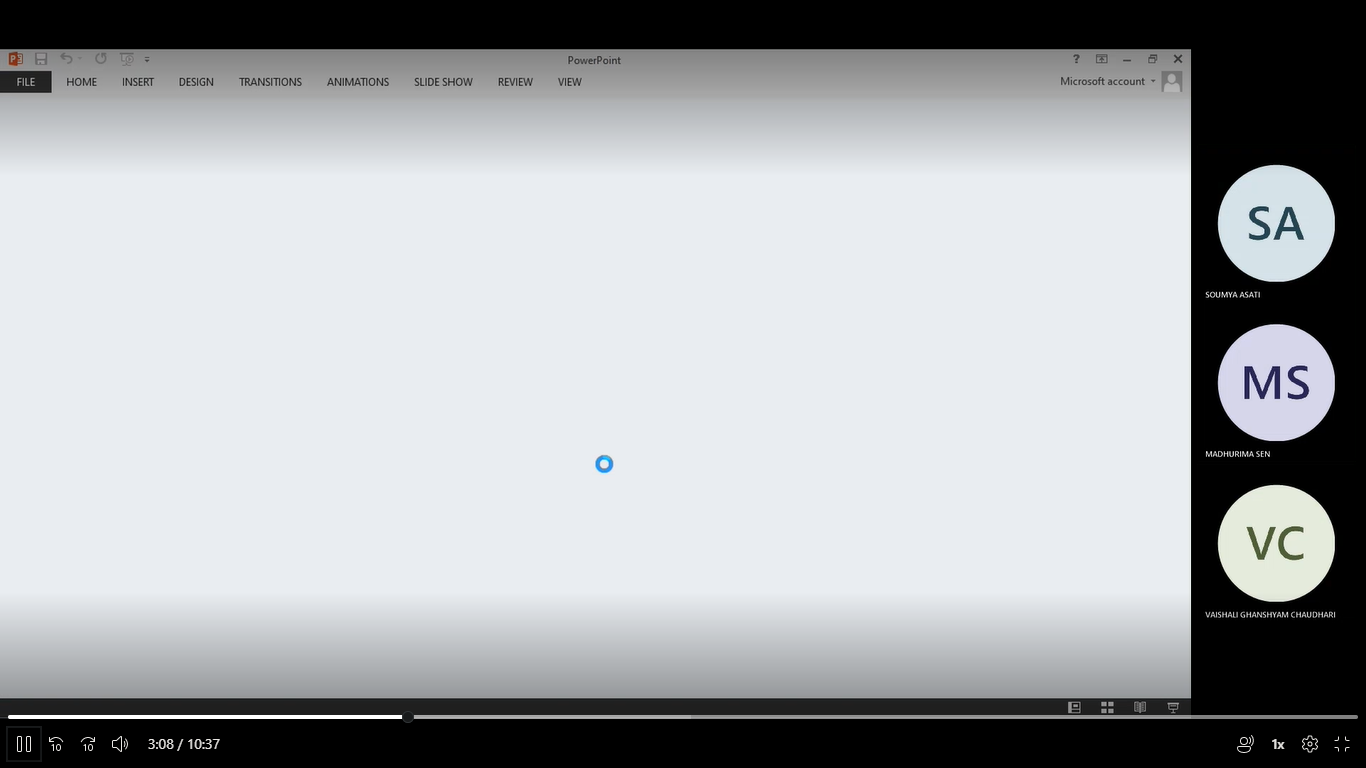


Fig. After open is detected, ppt file is opening

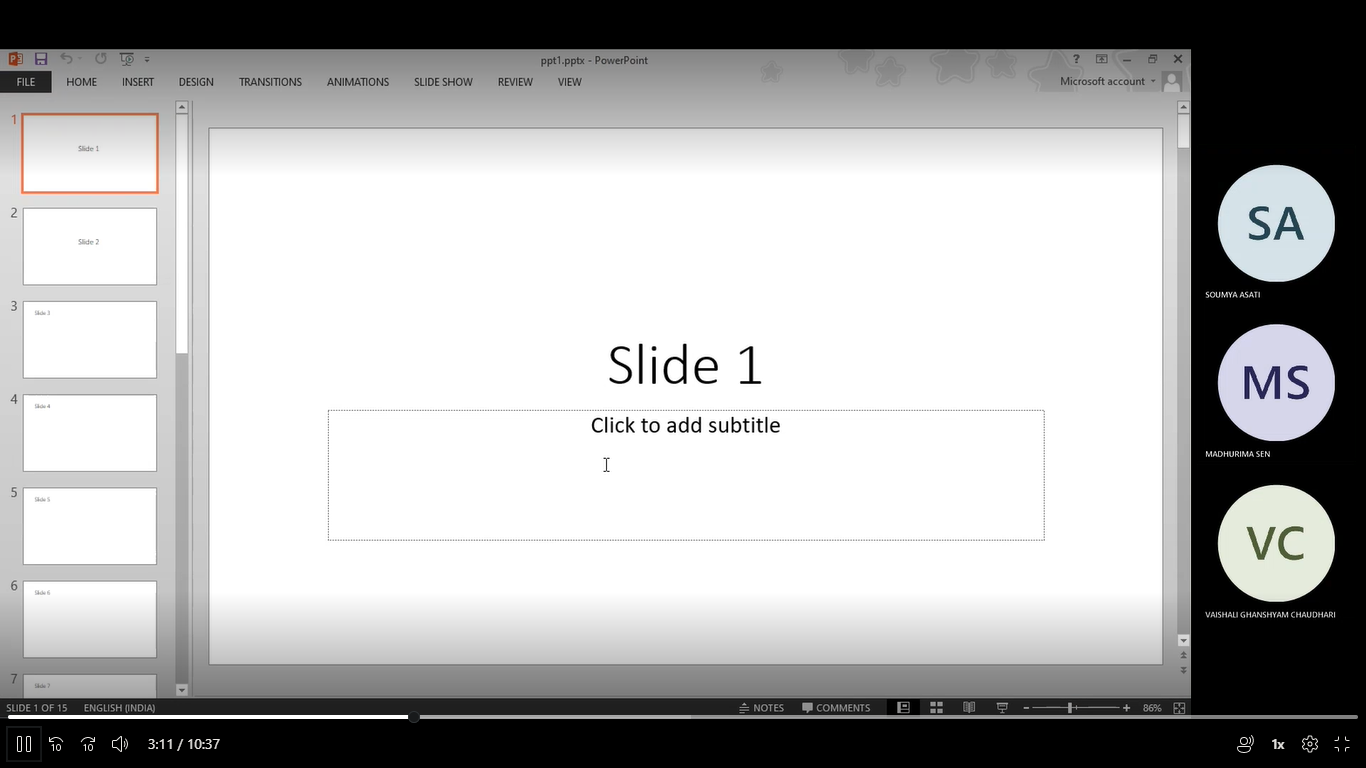


Fig. PPT opened after recognition of “open” word

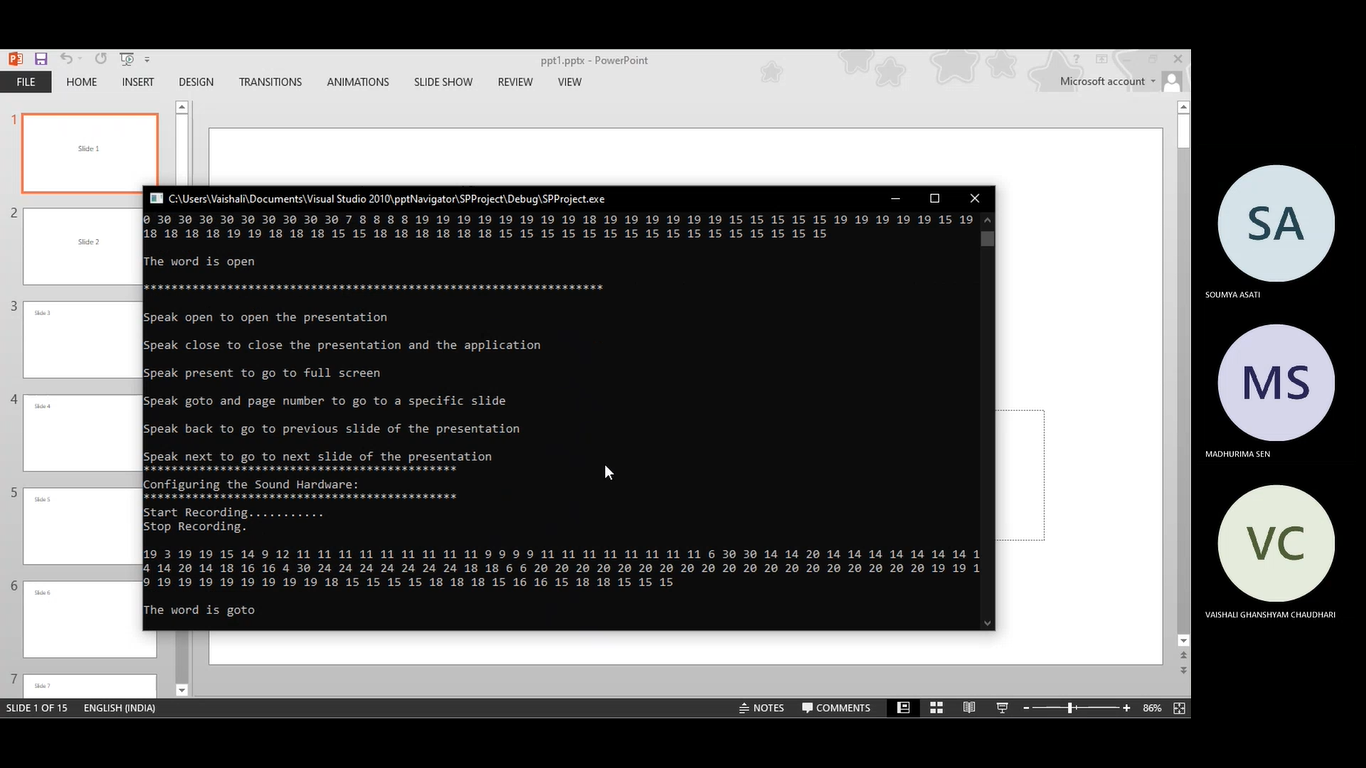


Fig. Goto word is recognized

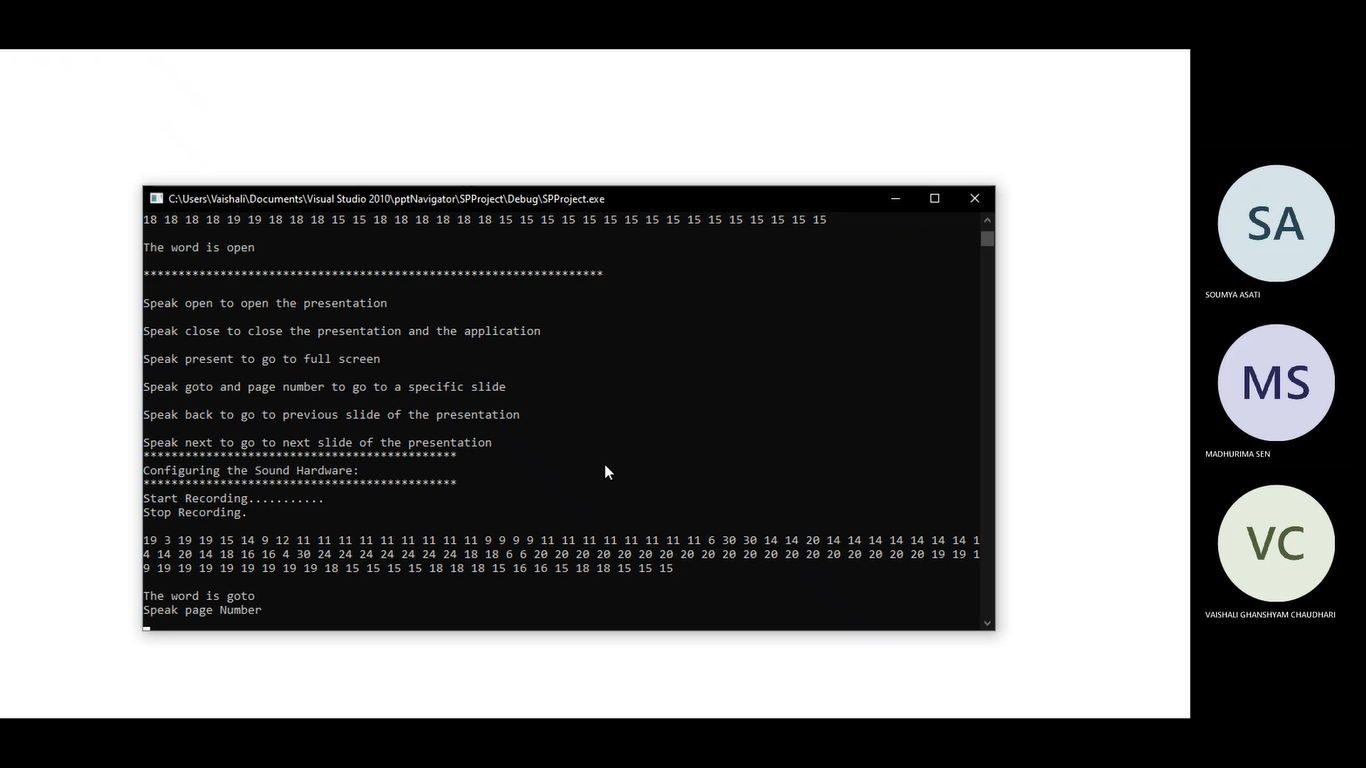


Fig. After goto word is recognized , Application asks for page number to goto

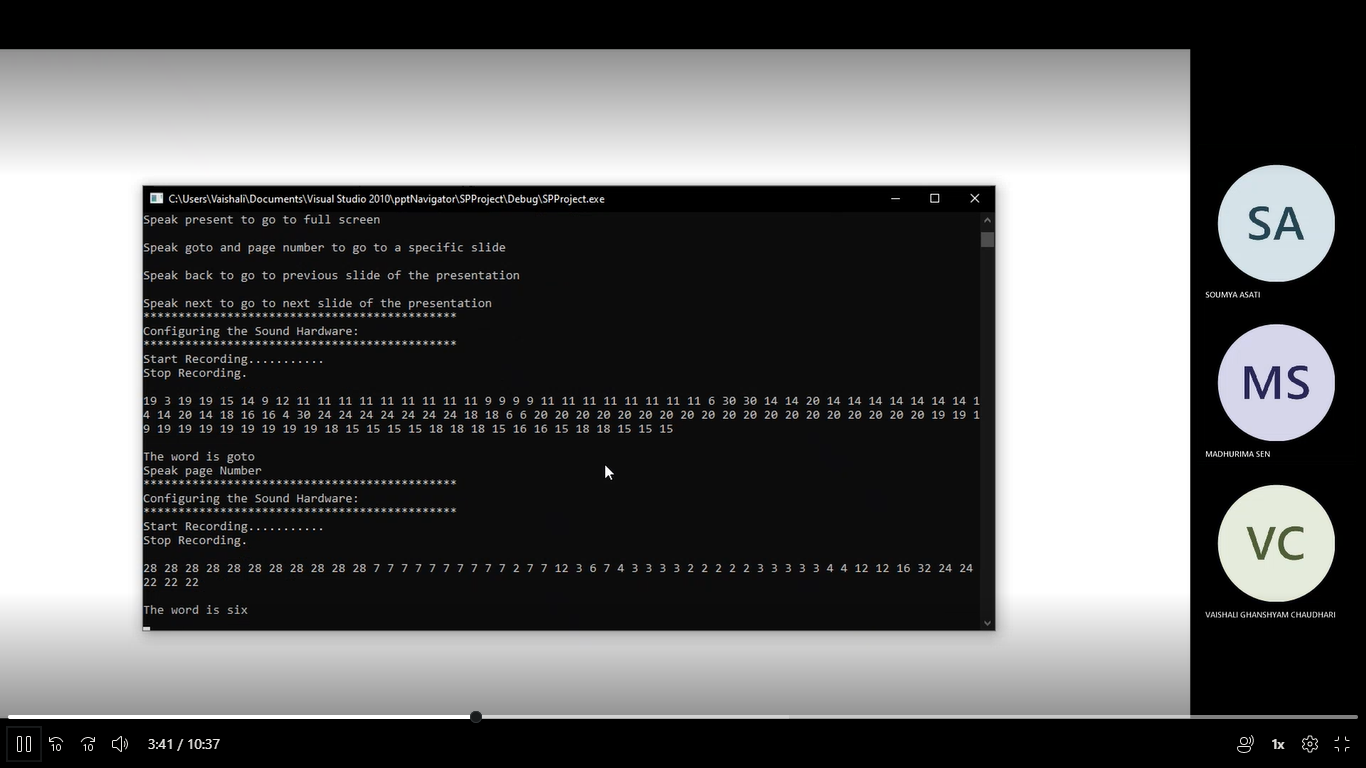


Fig. Page number “six” is recognized



Fig. slide number 6 has opened

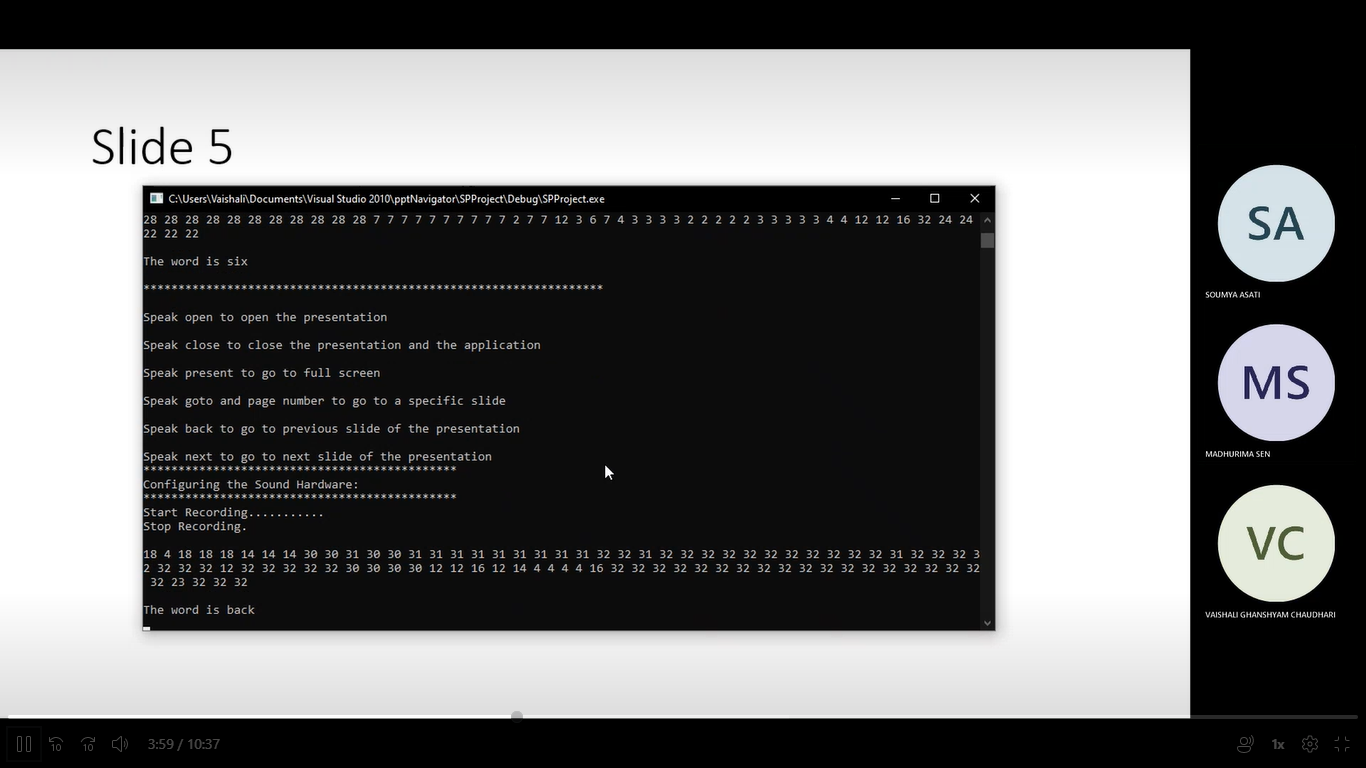


Fig. the word “Back “ is recognized



Fig. The slide goes to previous slide

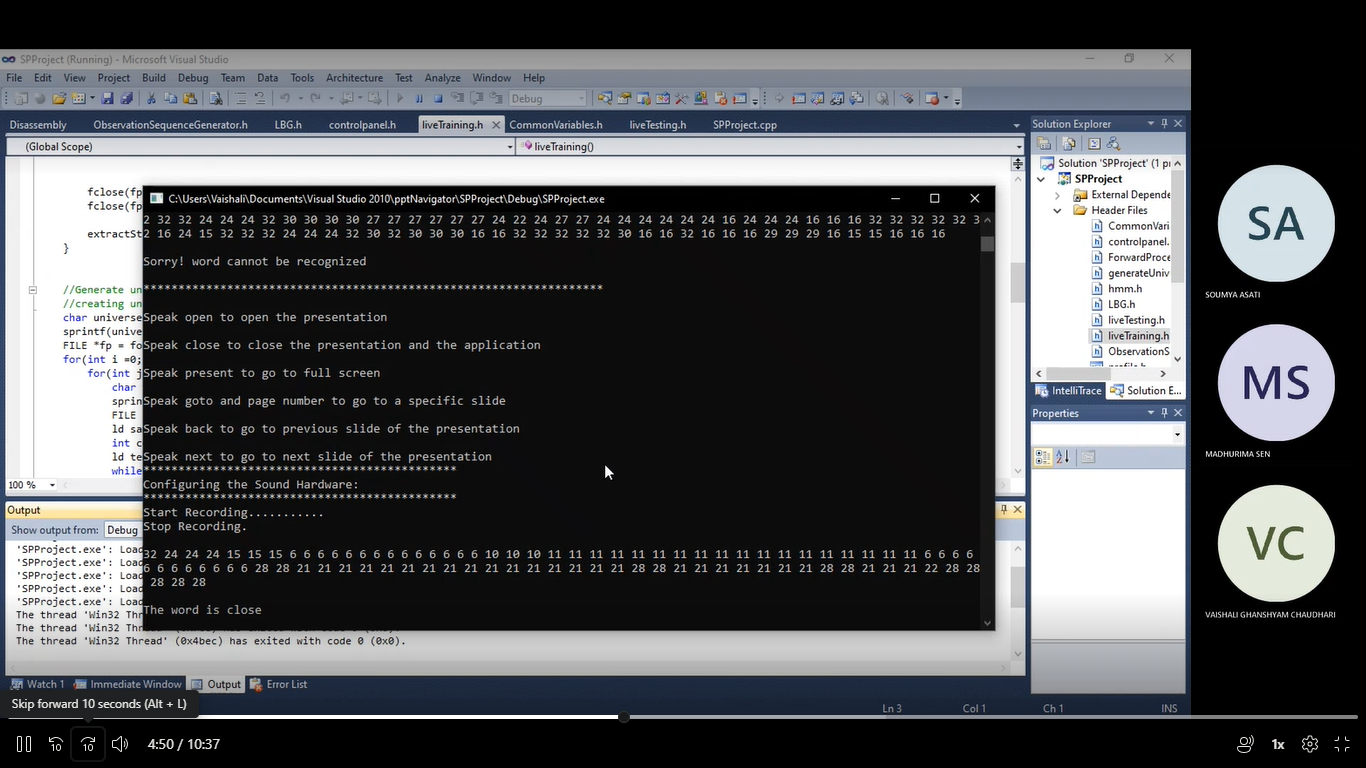


Fig. the Word “close” is recognized and PPT is closed.