Indian Institute of Technology, Guwahati

Department of Computer Science and Engineering

Project report

On

"Speech based Control Panel"

Based on

Speech recognition system

Course: CS566 Speech Processing

Submitted to

Prof. P. K. Das

Submitted by:

Paila Mouli Swaroop (214101035)
Senapathi Akhila Preethi (214101050)

Introduction:

We created a speech based voice assistant named "**SWAP**" which can open the following applications through the voice commands.

Applications listed:

1. Browser
2. Face book
3. Photos
4. Paint
5. Excel
6. Word
7. Notepad
8. Music

Proposed methodology:

Basic requirements to develop this project are as follows:

- Windows OS
- Microsoft Visual Studio 2010
- C++ 11 integrated with VS2010
- Recording Module

With the availability of above software, we further proceed in modelling the logic. The prerequisites of this project are

- Basic i/o operations on file
- Pre-processing of speech data
- Feature extraction
- Modelling of extracted feature
- Enhancing mode

Model Used for system development:

Hidden Markov Model:

When we cannot observe the state themselves but only the result of some probability function (observation) of the states we utilize HMM.

Elements:

1. Number of states (N considered to be 5)
2. Number of Observation symbols (M considered to be 32)
3. A, B and π matrices

Recordings:

- 20 utterances of each word are considered for the training and models are created.
- The recordings are done using cool edit software with sampling rate 16000 and 16bit mono channel.
- For live recordings, executable file "Recording_Module.exe" is used.
  Usage: Recording_Module.exe 3 input_file.wav input_file.txt

Codebook:

Cepstral coefficients of the frames considered in the training recordings are taken into the universe file for the generation of the codebook. A 32*12 size codebook is generated from the universe created using the LBG method.

Training Model:

The training was done with 20 utterances of each word. The frame limit was set to 150 at max. The average of model was done for three iterations, taking model generated at previous iteration as base. The threshold value of 10 to the power of -30 (difference in P* in consecutive iteration) was used to train the model for particular observation sequence.

Steps for training:

1. Pre-recorded utterances of words are used to generate the observation sequences.
2. The obtained observation sequences are used to train the models for different words.
3. The trained models are then used for testing to determine accuracy of the system.

Functionalities SWAP can perform:

Live Training:

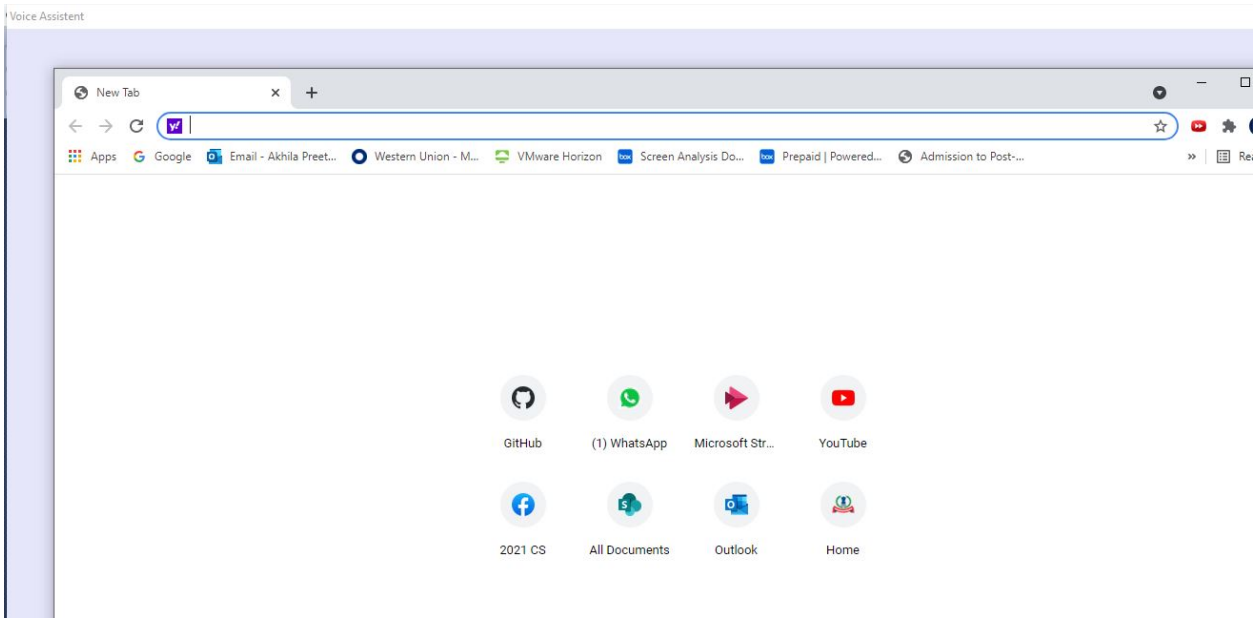
- For a new user we have included a module to train on user's voice.
- The user is prompted to utter the same word for 10 times.
- Data stored in the folder "live_recordings\\"
- New model is created for the word.
- Use the generated model for testing on the new user's voice.

Live Testing:

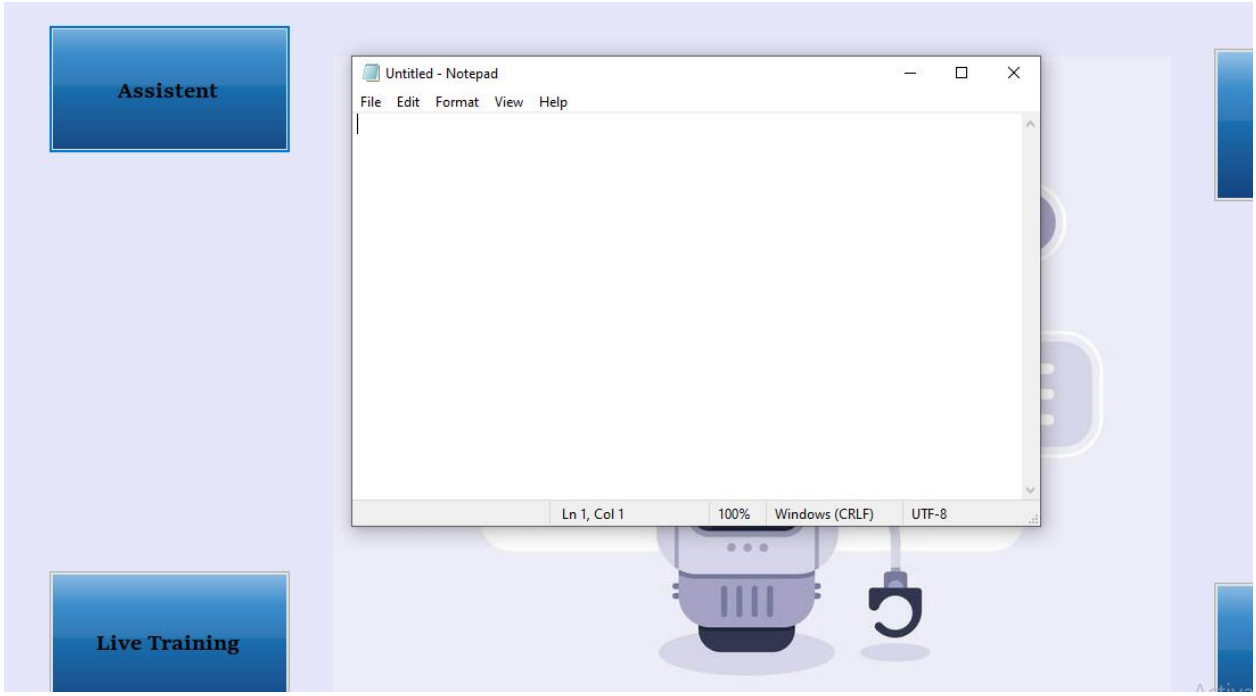
- SWAP prompts the user for the command (Application to be opened) .
- Through HMM models already trained, SWAP recognizes the command.
- SWAP opens the corresponding application.

User prompts and corresponding action by SWAP :

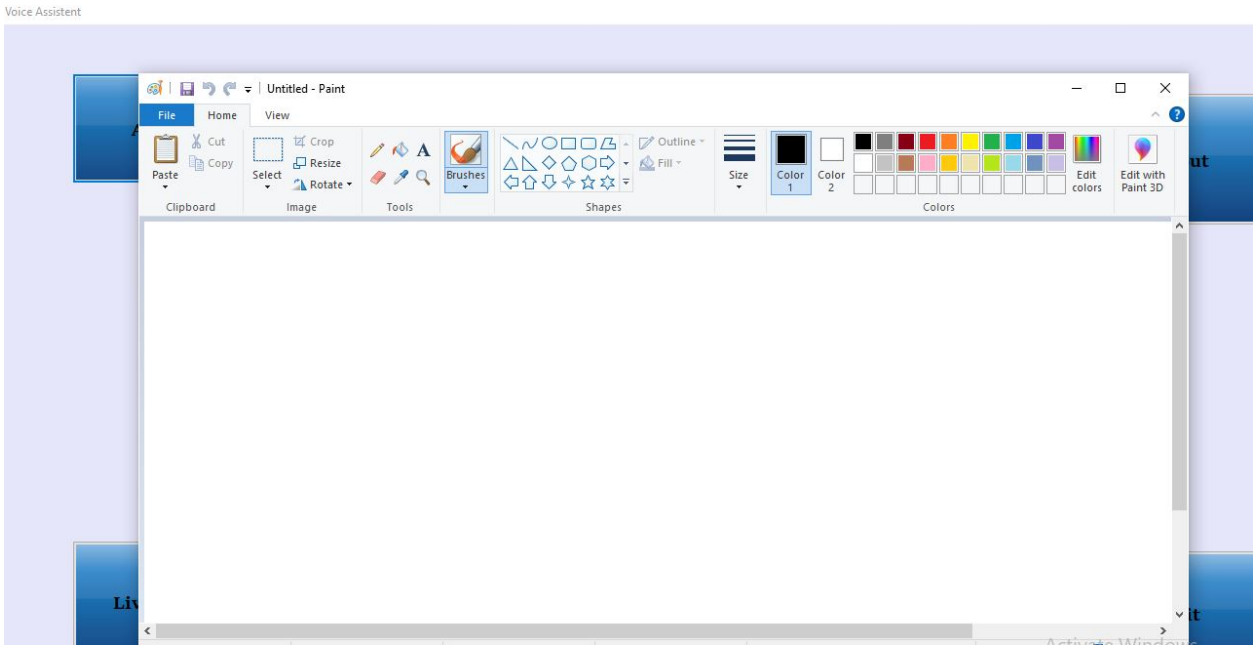
- **Browser**



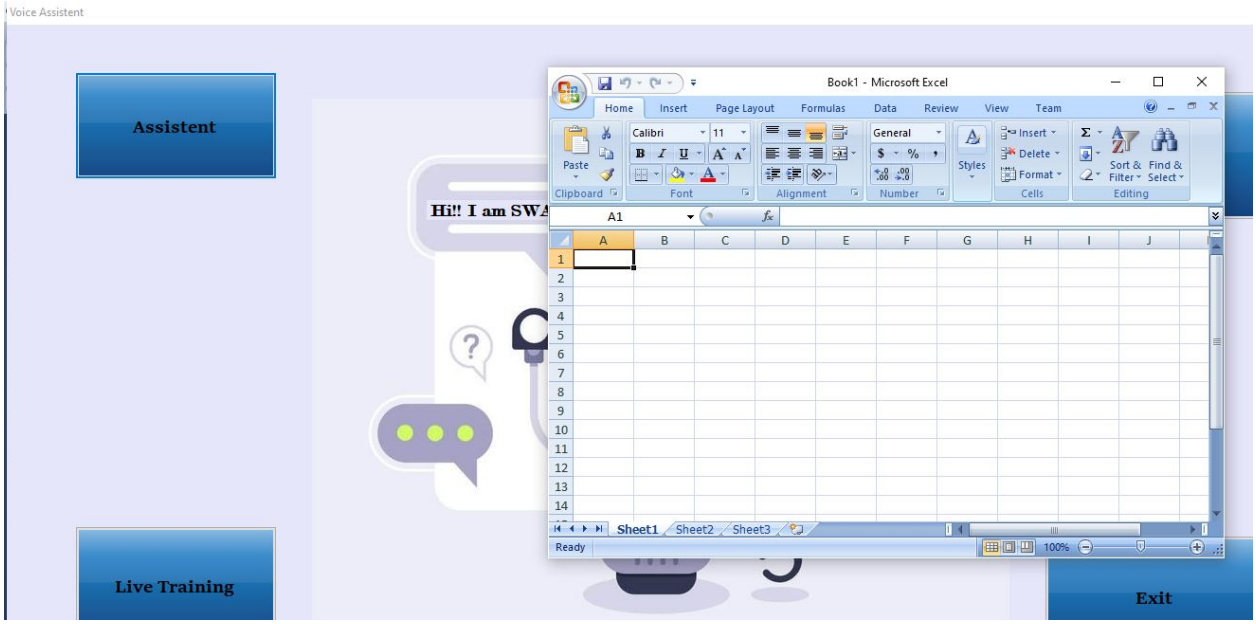
- **Notepad**



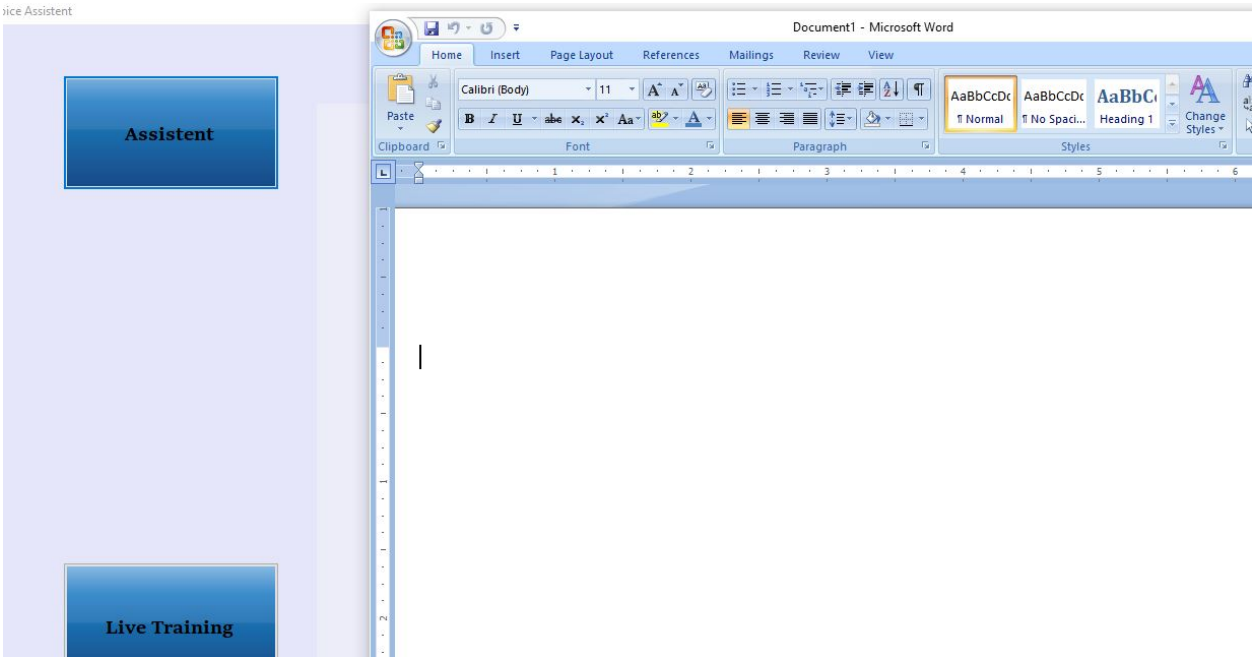
- **Paint**


Voice Assistent
Untitled - Paint
File
Home
View
Paste
Cut
Copy
Select
Crop
Resize
Rotate
Brushes
Outline
Fill
Size
Color 1
Color 2
Edit colors
Edit with Paint 3D
Clipboard
Image
Tools
Shapes
Colors

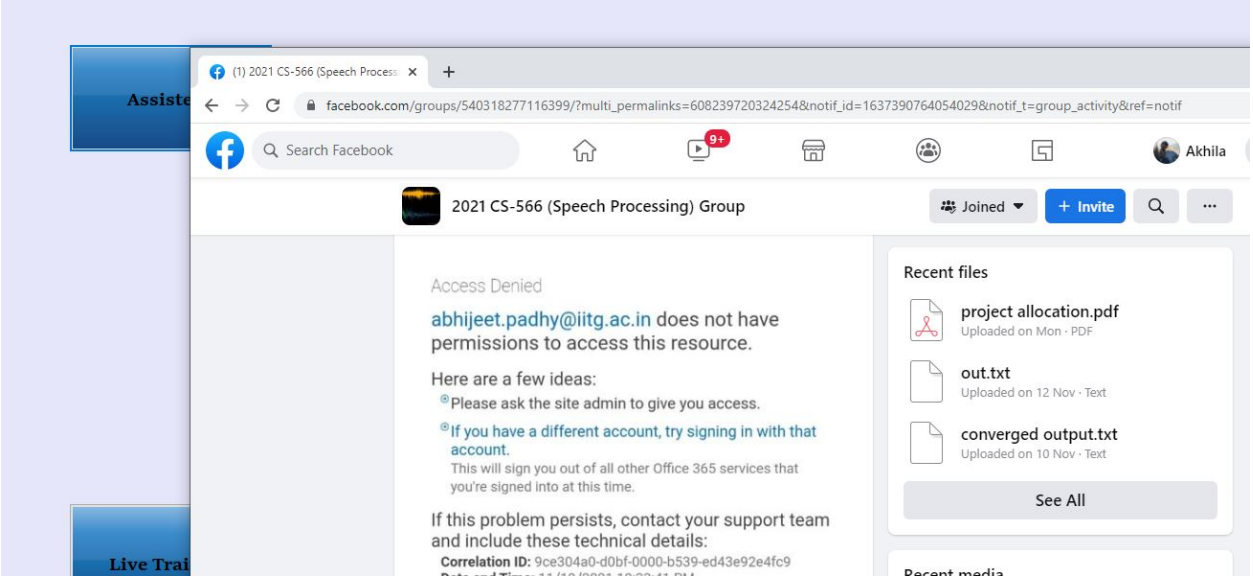
- **Excel**


Voice Assistent
Assistent
Hi!! I am SWA
Live Training
Exit
Book1 - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Team
Paste
Calibri
11
General
Styles
Insert
Delete
Format
Sort & Filter
Find & Select
Clipboard
Font
Alignment
Number
Cells
Editing
A1
Sheet1
Sheet2
Sheet3
Ready
100%

- **Word**



- **Face book**



- **Photos**

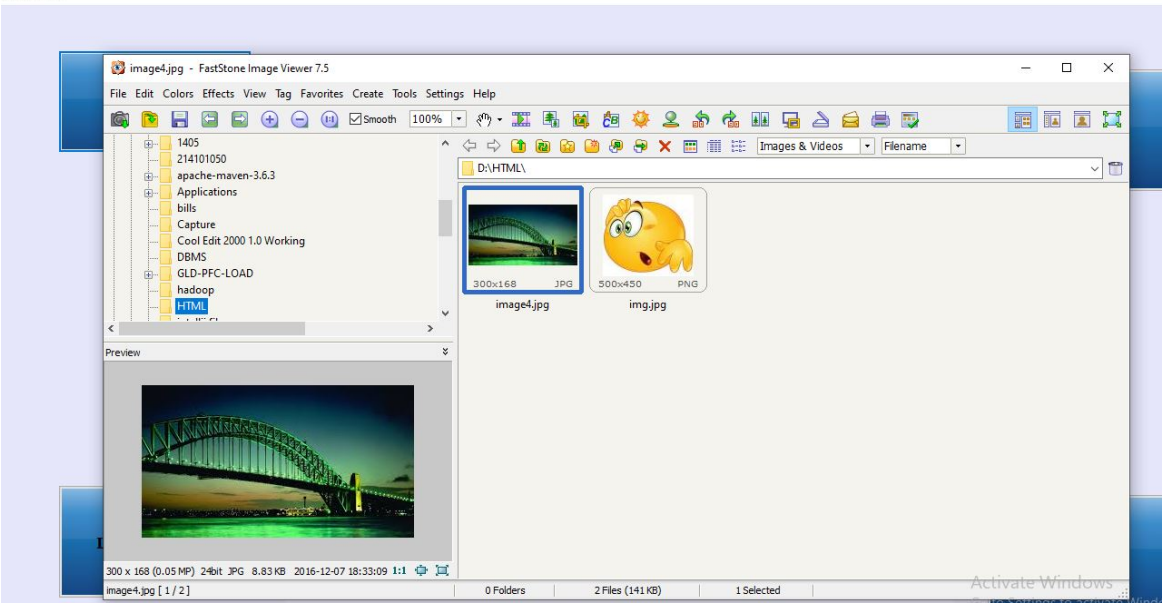


image4.jpg - FastStone Image Viewer 7.5
File Edit Colors Effects View Tag Favorites Create Tools Settings Help
Smooth
100%
Images & Videos
Filename
D:\HTML\
1405
214101050
apache-maven-3.6.3
Applications
bills
Capture
Cool Edit 2000 1.0 Working
DBMS
GLD-PFC-LOAD
hadoop
HTML
300x168 JPG
image4.jpg
500x450 PNG
img.jpg
Preview
300 x 168 (0.05 MP) 24bit JPG 8.83 KB 2016-12-07 18:33:09 1:1
image4.jpg [1 / 2]
0 Folders
2 Files (141 KB)
1 Selected
Activate Windows