Project Proposal
“Voice Command Automation"

KIET Group of Institution,Ghaziabad

(Affiliated to AKTU, Lucknow

Session 2023-24)

TABLE OF CONTENTS

1. Introduction

Voice automation is an emerging technology that has the potential to revolutionize the way we interact with our devices and surroundings. It enables users to control various appliances and devices using voice commands, eliminating the need for physical buttons or touchscreens. This project proposes the development of a voice automation system that can be used in both homes and offices, making daily tasks more convenient and efficient.
Our vision for the voice automation project is to create a system that seamlessly integrates into users' daily lives, providing them with a convenient and efficient way to control their environment using natural language.

2. Background

2.1 **History**

Our team consists of three members- Diya Bhatia, Prashant Kumar Singh and Shubham Bhatt. All of us are excellent learners and efficient workers . All of us have created some interesting projects in our respective fields and keeping in mind our unique skills and combined potential we have created this team.
Diya has worked on an app known as Movify which was a movie streaming app. It lets you and other members join a room, and using YouTube API you can stream a movie which every member can watch together.
Prashant and Shubham have worked together on an Optical Character Recognition System with TensorFlow and Keras API, which takes the image from the user and processes it into the computer readable format. Prashant’s major contribution was Object detection using OpenCV and Shubham’s major contribution was on Neural Networks and image processing using TensorFlow.

2.2 **Requirements**

Business Problem:

- Voice automation can greatly benefit individuals with disabilities or limited mobility, allowing them to control their environment without relying on physical buttons or touchscreens.
- Voice automation can also improve efficiency by eliminating the need for users to physically interact with their devices.
- Voice automation can also improve security by enabling users to control their security systems using natural language.
- Overall, there are many problems that could benefit from the implementation of voice automation technology. By enabling users to control their environment using natural

language, voice automation can improve accessibility, efficiency, safety, and security in a variety of settings.

Business opportunity:

- Smart Homes: Voice automation can enable users to control their home appliances, lighting, temperature, and security systems using natural language. This can make daily tasks more convenient and efficient, while also improving energy efficiency and home security.
- Healthcare: Voice automation can be useful in healthcare settings, allowing doctors and nurses to quickly and safely adjust lighting, temperature, and medical equipment without having to touch any buttons or equipment.
- Retail: Voice automation can be used in retail settings to improve customer experience and streamline operations.
- Automotive: Voice automation can be used in cars to improve driver safety and convenience.
- Manufacturing: Voice automation can be used in manufacturing settings to improve worker safety and efficiency.

2.3 **Solution**

Components Used:

- Microphones: Microphones are used to capture audio input from the user. They are typically placed in strategic locations within the environment to ensure that they can pick up the user's voice clearly.
- Neural Networks: System must be able to recognize the designated keywords that are assigned to it, in order to perform the correct task. This will include, dataset, preprocessing of sample voice, Convolution Neural Networks, User Input, etc.
- Integration with Other Systems: Arduino's powerful features are a good fit for home automation. There is a large scope of home automation applications that we can build by using the Arduino board. Voice automation systems must be able to integrate with other devices and systems in the environment to provide full functionality. This may include integration with smart home devices, security systems, or other third-party applications.

Assumption:

- Speech clarity: Voice recognition models assume that the speech input will be clear and easy to understand. If the audio input is muffled, distorted, or contains background noise, the model may have difficulty interpreting the user's command.
- Vocabulary limitations: Voice recognition models have a limited vocabulary and can only recognize words and phrases that are included in their language model. If the user speaks a word or phrase that is not in the model's vocabulary, it may not be recognized.

- Accent and dialect variations: Voice recognition models are trained on a particular dialect or accent and may have difficulty understanding users with different speech patterns. Models can be fine-tuned to better understand different accents, but they may still have some limitations.
- Homophones and similar sounding words: Voice recognition models may have difficulty distinguishing between words that sound similar or are homophones (e.g. "there" and "their").

In-Scope:
The application of Voice Automation is ever-expanding. It is used to control home appliances. It is highly useful for disabled people who may not be able to physically interact with devices and perform actions. It can be used to voice automate the devices such as light, fan, plugs, etc.

Out-of-Scope:

- With the current model only basic operations can be performed . These include turning on and off of devices like fans , lights , speakers and AC's. We cannot perform complex operations which require highly specific commands for example , regulating the speed of fans, changing channels, we can only operate either on or off feature.
- Currently, our model works on a limited number of devices. In order to add a new device, the model needs to be trained again, for the particular keyword.

Dependencies:
To create a speech recognition system using TensorFlow, one can follow the below steps:

1. Download .wav files: You can download .wav files from Google research blog or any other source. These files will be used to train the speech recognition model.

2. Preprocess the audio files: Before feeding the audio files to the model, you need to preprocess them. This involves decoding the audio file, binarizing it, creating spectrograms, and converting it into a format that can be fed to the model.

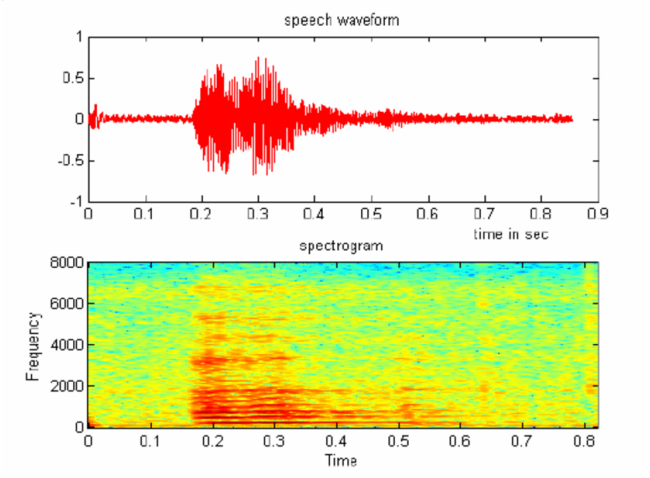
3. Build a Convolutional Neural Network (CNN) model: TensorFlow provides APIs to build a CNN model. You can use these APIs to create a model that takes the spectrogram as input and classifies it into one of the predefined keywords. You can use the softmax function to output probabilities for each keyword.

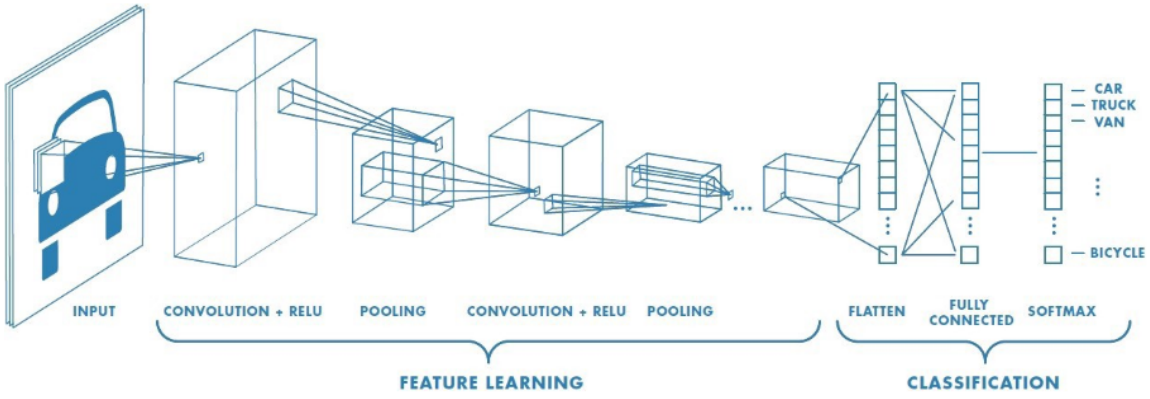
4. Train the model: Once the model is built, you can train it using the preprocessed audio files. TensorFlow provides APIs to train the model using the training data.

5. Test the model: After the model is trained, you can test it using the test data. You can measure the accuracy of the model by comparing the predicted output with the actual output.

6. Integrate with home appliances: Once the model is tested and you are satisfied with the accuracy, you can integrate it with home appliances. This can be done by connecting the model to a microcontroller that can control the home appliances based on the predicted output.

Overall, creating a speech recognition system using TensorFlow involves preprocessing the audio files, building a CNN model, training and testing the model, and integrating it with home appliances. With the help of this system, you can create a smart environment where you can control home appliances using voice commands.





3. Proposal

3.1 **Deliverables**

Project Deliverables	
Title	Description
1. Recognition of Audio Commands	Keywords when spoken are recognized by the model.
2. Automation of Lights and Fans	Using voice commands to turn on and off lights and fans.

3.2 **Timeframe**

Initiation- Coming up with Project Idea 2 days
Planning - Formulating Project Plan 7 days
Executing - Creating the Project 65 days

3.3 **Resources**

- Hardware
  - Computer
    - Central Processing Unit (CPU) — Intel Core i5 6th Generation processor or higher. An AMD equivalent processor will also be optimal.
    - RAM — 8 GB minimum, 16 GB or higher is recommended.
    - Graphics Processing Unit (GPU) — NVIDIA GeForce GTX 960 or higher. AMD GPUs are not able to perform deep learning regardless.
    - Operating System — Ubuntu or Microsoft Windows 10.
  - Microphone
  - Arduino Board

- Software
  - Python 3
  - Libraries - Tensorflow , OpenCV, Pandas, Numpy

3.4 **Risks & Issues**

Project Risks		
Risk	**Details**	**Likelihood**
Unavailability of Hardware	If the availability of hardware is delayed or denied	Low
System Requirements	If the system is not capable of processing the data	Medium
Integration with Devices	Implementing the software for day-to-day appliances requires an additional hardware.	High

Project Issues		
Issue	**Details**	**Impact**
Unavailability of Dataset	Proper dataset with appropriate keywords couldn't be found.	Medium

4. Appendix

4.1 **Supporting Documentation**

Links and References:

1. DeepLearning.AI TensorFlow Developer Professional Certificate | Coursera
2. Build your own real-time voice command recognition model with Te…
3. Google Dataset