A VOICE INTERFACE FOR CONTROLLING HARDWARE DEVICES

PROJECT REFERENCE NO.: 39S_BE_1491

COLLEGE : NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL
BRANCH : DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GUIDE : DR. SHASHIDHAR G KOOLAGUDI
STUDENTS : MR AVINASH B M
MR SHIVA SAGAR B
MR SHREYAS UDUPA B

INTRODUCTION:

With the advent of speech recognition technology, there is a lot of scope for voice activated devices. It is a viable option for multitasking. With voice activated devices, consumers can handle multiple tasks and appliances with ease as there would be no need to operate buttons, switches and dials for each device. Applications for such a system include home automation possibilities, aircraft control, educational systems and robot control. These systems would also be very beneficial for people with disabilities. Design of voice activation systems primarily involves work on speech recognition and hardware interface.

OBJECTIVE:

The aim is to develop an easy-to-use and affordable speech interface for hardware devices. The system involves a composition of an automatic speech recognition system (speech – to – text converter) and a hardware interface which utilizes a microcontroller. The automatic speech recognition system shall be designed for a customizable vocabulary consisting of eight commands. The system shall be tuned to be environment – independent by removing unvoiced portions of the speech and other background noises. The system shall use a set of 8 LEDs on the output side as representatives of common everyday devices. The input will be from a human speaker through a standard microphone. The user will be able to activate or deactivate individual LEDs with simple spoken commands.

The following are the objectives of the project:

- Robust Speech Recognition System to recognize spoken commands
- Using serial communication to transfer results to a microcontroller
- Developing a control logic at the microcontroller to toggle devices

METHODOLOGY:

The entire system is a combination of two systems, a speech recognition system and a hardware system comprising of LEDs which are representative of a customizable set of hardware devices. The two systems are bound together in operation by a control logic coded into an Arduino Uno, a micro-controller. The speech recognition system is a combination of a
signal processor and a machine learning algorithm involving a classifier. The communication between the speech processing software and the micro-controller is via a serial port.

A set of words is chosen by the user as a command set. The user's voice is used to train the speech recognition system. During usage, the user speaks the intended command into a microphone. The input is passed onto the speech recognition system. Here, the audio input is first preprocessed to remove noise. A trained classifier is then used to recognize the command. Thus, the command is obtained in textual format. This input is converted into a decimal number and fed into the micro-controller via a serial communication port. This acts as input to the control logic which causes the intended output port to receive high voltage in order to activate the devices which are to be controlled.

**Feature Extraction** - The voicebox library in MATLAB is used for this purpose. As feature vectors, Mel Frequency Cepstral coefficients are derived from the speech signal and 16 coefficients obtained as amplitudes of the resulting spectrum are used.

**Speech Recognition** - In this system, one vs. all logistic regression is implemented to classify between multiple spoken commands. Since the system in use has eight different commands, eight different logistic regression classifiers are trained. The audio is recorded and initial filters are applied to remove noise and unvoiced regions. The features used here classification using logistic regression are sixteen coefficients of the Mel frequency cepstrum. If there is incompatibility between multiple datasets, padding is used with zeros to normalize all feature vectors to a constant length of two hundred. Using these feature vectors, logistic regression classifiers are trained with regularization to avoid over-fitting as illustrated.

**Microcontroller** - Once the spoken command is recognized and converted into textual format, the next step is to convert it into output on a hardware device. A microcontroller is used to interface the software with the hardware being used. In the system developed, Arduino Uno is used as the microcontroller. The microcontroller provides digital and analog input and output pins that can be used for interfacing with electronic circuitry. It also features serial communication interfaces including Universal Serial Bus which can be used to load executable programs from personal computers. Using a dictionary of commands built from the pre-defined command set, the recognized command is converted into a number. In this system, it is a number between one and eight. A program is saved and run on the microcontroller which acts as the control logic. The number obtained after indexing is fed as input to this program. Digital pins five through twelve are designated as output pins for the system. Based on the number received, one of the output voltages is set high. This turns on the device connected at the pin. If previously, the device was on, the device is then switched off. Each command essentially performs a toggle function on the corresponding output device.

**Hardware Devices** - Devices are connected at the output side of the microcontroller which can be switched on or off with single voice commands. In the system developed, for demonstration purposes, a series of Light Emitting Diodes are placed on a breadboard to simulate everyday appliances. The states of these LEDs, on or off, are controlled using voice commands only. The LED array on the breadboard consists of eight LEDs. The negative nodes of all these LEDs together constitute a common grounding point. This is connected to the pin labelled ground on the microcontroller. The positive end of each LED is connected to the corresponding output pin of the microcontroller via a resistor and a wire. The resistor ensures that the voltage received by the LED is sufficiently high when activated to turn it on and not too high as to burn it out. The LEDs used at the output side are representative of the devices that could actually be connected and controlled using human voice.
CONCLUSION:

An efficient and cost-effective system is created which performs as an easy-to-use voice interface to control hardware devices. This system is customizable. The command set can be chosen by the user. The user will have to train the system to recognize the set he chooses for the commands.

The speech recognition system is robust and achieves an accuracy of at least 90%. One of the observations is that for an isolated word recognition task, One versus All Logistic Regression performs better than Hidden Markov Models by obtaining an accuracy that is better by seven percentage points. Training logistic regression models is also much faster. Logistic regression models can be trained in less time compared to Hidden Markov Models.

FUTUREWORK:

The following improvements can be made to the system to obtain improved results and performance. The system should have a flexible vocabulary. The current system has a set of commands that is predefined. Although this set is customizable, there is not much flexibility with the commands themselves. If 'air conditioner' is one of the predefined commands, the system should be in a position to also recognize 'AC' and 'cooler' as the same command. The system should be able to pick commands from continuous speech. Instead of just picking discrete utterances, it should be able to segregate appropriate commands from continuous speech. An accuracy of 92% is achieved with logistic regression. With the use of neural networks and more training data, the recognition accuracy can be increased. The system lights up LEDs on the output side for demonstration purposes. With the aid of relevant circuitry, the microcontroller output can be connected to appliances on the power grid. The system can then be used efficiently in room-like environments.