AI-Based Speech Recognition System for Smart Devices

Sai Kiran Bagli
Dept. of ECE
MLR Institute of Technology
Hyderabad, India
Kiransai2589@gmail.com

Satya Prakash Dash
Dept. of ECE
MLR Institute of Technology
Hyderabad , India
Satyaprakashdash879@gmail.com

Shashanth Bhaidishetty
Dept. of ECE
MLR Institute of Technology
Hyderabad, India
Bhaidishettyshashanth@gmail.com

Pradeep Oalati
Dept. of ECE
MLR Institute of Technology
Hyderabad, India
19R21A0497@mlrinstitutions.ac.in

Abstract*—Automatic Speech Recognition (ASR) has progressed considerably over the past several decades, but still has not achieved the potential imagined at its very beginning. Almost all of the existing applications of ASR systems are PC based. This thesis is an attempt to develop a speech recognition system that is independent of any PC support and is small enough in size tobe used in a daily use consumer appliance. The main objective is to develop a low-cost home automation system using AI based ASR system which is easy to install and configure & to embed a speech control interface for controlling the electrical devices. In this project, the advanced version of developed (to be developed by us) ASR technologies are to deploy into hardware for estimating their real time performance. This hardware was trained and tested with certain commands to perform various tasks. It also saves human time in daily life as it is automated, and user can also operate it by sitting at one place as we added voice recognition feature to our model.** ***Keywords—Artificial Intelligence (AI), Automatic Speech Recognition (ASR)

I. INTRODUCTION

Technology has increased the rate at which data is processed by machines in our lives. These machines include computers, televisions, microwave ovens and many others. As the data processing rate increases, the machines begin to wait for human data input, rather than humans waiting for the machines to respond. There are many input technologies available today that include keypad units (Keyboards, remote controls, or membrane switch matrices) and pointing devices (mice, joysticks, or digitizing tablets).[1] For humans, speech plays an important role to communicate efficiently. Several languages are used for communication in different parts of the world. In recent years, with advancement in machine learning technologies, the demand of Automatic Speech Recognition Systems has risen. ASR systems have basically three steps: Acquisition of speech signal, Feature extraction and Pattern matching.

Speech Recognition is the technology that allows human beings to use their voices to speak with a computer interface & converts human speech into readable text. Speech recognition with AI means adopting digital assistants, automated support & human – robot interactions in the form of voice user interfaces (VUIs) to streamline their services. Smart devices are all the everyday objects made intelligent with advanced compute, including AI and machine learning, and networked to form the internet of things (IoT). Smart devices can operate at the edge of the network or on very small endpoints, and while they may be small, they are powerful enough to process data without having to report back into the cloud. Smart home devices leverage speech recognition technology to carry out household tasks, such as turning on the lights, operating fan, boiling water, adjusting the thermostat, and more.

II. METHODOLOGY

Implementation of ASR :

Training the model by acoustic, lexicon models. Testing the model by language model. At the end performance is evaluated by decoding. Implementation in Hardware(IOT) :- In this project, the advanced version of developed (to be developed by us) ASR technologies are to deploy into hardware for estimating their real time performance.

A neural network approach for classification using features extracted by a mapping is presented. When the number of sample dimensions is much larger than the number of classes and no deviations are given but the means of classes, a mapping from class space to a new one whose dimensions is exactly equal to the number of classes is proposed. The vectors in the new space are considered as the feature vectors to be inputted to a neural network for classification. The property that the mapping does not change the separability of the original classification problem is given. Simulation results for speech recognition are presented.

A. Feature Extraction

It is the fundamental step in the Automatic Speech Recognition (ASR) process. In this pre-processing relevant data are extracted from a speech while removing background noise and irrelevant information.Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data.

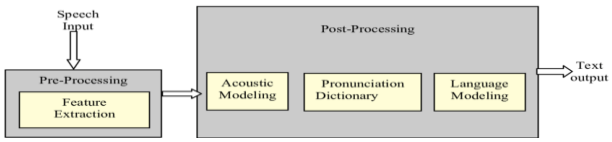

Fig.1. Feature Extraction

B. Acoustic Model

The acoustic model (AM), models the acoustic patterns of speech & maps raw data feature to phonemes based on accent. The job of the acoustic model is to predict which sound or phoneme is being spoken at each speech segment from the forced aligned data by HMM or GMM.Acoustic modelling of speech typically refers to the process of establishing statistical representations for the feature vector sequences computed from the speech waveform. Hidden Markov Model (HMM) is one most common type of acoustic models. Other acoustic models include segmental models, super-segmental models (including hidden dynamic models), neural networks, maximum entropy models, and (hidden) conditional random fields, etc.

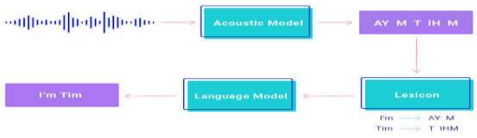

Fig.2. Acoustic Model

C. Lexicon Model

This is phoneme to word mapping. This model contains pronunciation models which describes how words are pronounced phonetically. The reason why the lexicon is such an important piece of the speech recognition pipeline is because it gives a way of discriminating between different pronunciations and spellings of words. For example, consider “ough” as in through, dough, cough, rough, bough, thorough, enough, etc. The pronunciation cannot be known from the spelling. In this case, the correct pronunciation for a given word will be determined contextually from the lexicon and the state transition probabilities which it encodes between word/phoneme states.

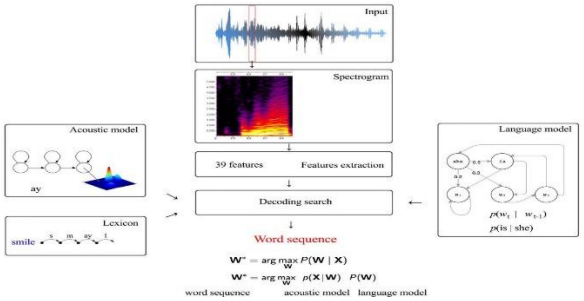

Fig.3. Lexicon Model

D. Language Model

This defines how words are connected to each other in a sentence. It learns which sequences of words are most likely to be spoken, and its job is to predict which words will follow on from the current words and with what probability & finally it produces output. They are used in natural language processing (NLP) applications, particularly ones that generate text as an output. Some of these applications include, machine translation and question answering.

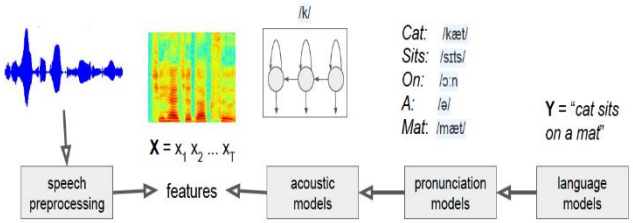

Fig.4. Language Model

III. PRINCIPLE

The working of ASR module is simple but the technology is impressive. The working is made possible with the help of Voice Recognition Module & Arduino UNO. The Recognition Module is trained and tested with the software Access Port to import commands in 3 – Groups of total 15 commands. Arduino UNO is programmed for working of lights, servo motor, speaker etc. The Recognition Module, it was trained by Access port software with the HEX values where,

- HEX AA 36 is used for common mode.
- HEX AA 11 is used for recording mode.
- HEX AA 21 is used for importing commands into module.

The Arduino UNO is programmed for operating servo motor, lights by trained commands.

A. Operation

First raw audio input is given through mic which is connected to voice recognition module and then connected to ttl module.

Voice commands recognized by Access Port software. In this step training and testing of system is done. After this as the commands are stored in voice recognition module. This is connected to Arduino which has code to perform appropriate actions such as turning on the lights, operating fan, rotating servo motor, and more.

The system is speaker independent and is suitable for real time processing. Each time the testing command is matched with the trained commands, the system includes the input speech data into its training data-set, thereby enriching the training database, for next comparisons. The use of Euclidean distance measurement and after that neural networks make the system two level secure.

Table 1. Components Price List

Component name	Quantity	Cost
Voice Recognition Module	1	1800
Arduino Uno	1	650
Bread Board	1	80
TTL module	1	50
Jumper Wires	-	60
LED'S	3	10
Fan	1	50
Servo motor	1	40
Total		2740

B. Block Diagram

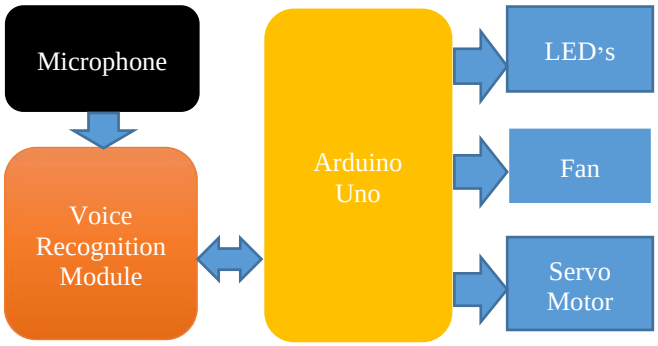

Fig.5. Block Diagram of Working

The Arduino UNO is central part or brain of the working process, as it is the main component for processing. The Arduino UNO R3 is the perfect board to get familiar with electronics and coding.. Another important component is Voice Recognition Module where the input is taken from microphone connected module.

C. Software

The Software we used Access Port is a freeware rs232 monitoring app that's been categorized by our editors under the programming software category and made available by WWW.SUDT for Windows. The review for Access Port has not been completed yet, but it was tested by an editor here on a PC and a list of features has been compiled; see below. We've also created some screenshots of the user interface to show the overall usage and features of this rs232 monitoring program.

III. DESIGN SETUP

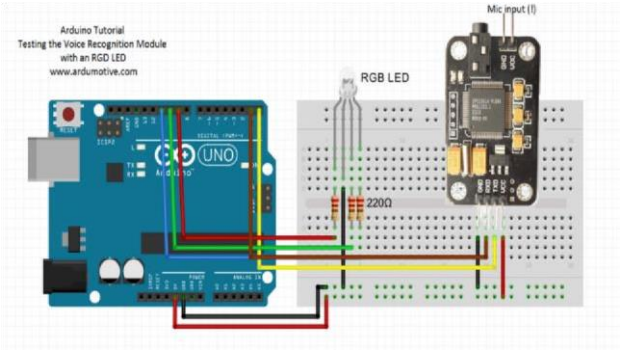

Fig.6. Circuit Diagram

IV. RESULTS AND DISCUSSIONS

Step 1:*Audio commands will be sent & captured through microphone.*

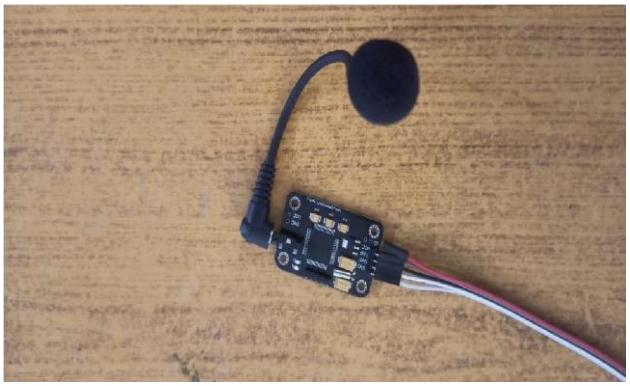

Fig. 7. Commands sending through microphone

Step 2:

- The recognized commands will be loaded into software using TTL module.
- In software by using HEX AA 21 the commands will be imported into voice recognition module.

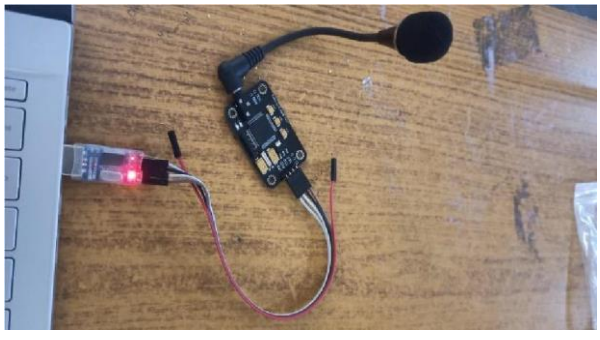

Fig. 8. Commands importing

Step 3:

- The Arduino will identify the commands and perform the appropriate action.
- The actions performed are turning on LED, Serve Motor, Speaker.

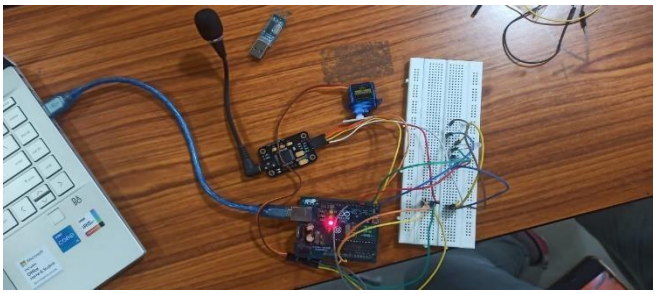

Fig. 9. Working of Prototype

V. CONCLUSION

It can be concluded that the design and implementation that has been done using voice commands in the form of voice recognition and speech recognition can facilitate human activities in everyday life. In addition, the application of voice commands can also be applied to smart rooms such as controlling lights, servo motors, speaker. Controlling the home utilities via voice is just an amazing step forward towards the development in IoT sector, as this totally involves a wireless medium to create the connection. There are many Android-based applications which have been developed to initiate the working on this technology which also includes voice commands. Without a doubt, this technology will bring revolution in the people's life if that is implemented on the larger scale. After performing deep research and study, we have introduced a platform, in which more efforts can result in the better format in future.

VI. ACKNOWLEDGMENT

We express our profound thanks to the management of MLR Institute of Technology, Dundigal, Hyderabad, for supporting us to complete this project.

We take immense pleasure in expressing our sincere thanks to Dr.K. Srinivasa Rao, Principal, MLR Institute of Technology, for his kind support and encouragement.

We are very much grateful to Dr S.V.S Prasad, Professor & Head of the Department, MLR Institute of Technology, for encouraging us with his valuable suggestions. We are very much grateful to Dr. Bittu Kumar, Assistant Professor for his unflinching cooperation throughout the project.

VII. REFERENCES

[1] M. R. Samburu, N. S. Jayant, "LPC analysis/synthesis from speech inputs containing quantizing noise or additive white noise", IEEE Trans. Acoustic.
Speech Signal Processing, vol. ASSP-24, pp. 488-494, Dec. 1976.

[2] Jihyuck Jo , Hoyoung Yoo and In-Cheol Park "Energy-Efficient FloatingPoint MFCC Extraction Architecture for Speech Recognition Systems", IEEE Transactions on Very Large Scale Integration (VLSI) Systems ,Volume: 24, Issue: 2,pp. 754 – 758, Feb. 2016 .

[3]RASTA Filtering", Proceeding of NationalConference on Communications Ram Singh, Preeti Rao, "Spectral Subtraction SpeechEnhancement with
(NCC), Kanpur, India, 2007.

[4] H. Doi, K. Nakamura, T. Toda, H. Saruwatari, K. Shikano, "Statistical approach to enhancing esophageal speech based on Gaussian mixture models," Proc. ICASSP2010, pp. 4250–4253 (2010).

[5] Myungjong Kim, Younggwan Kim, Joohong Yoo "Regularized Speaker Adaptation of KL HMM for Dysarthric Speech Recognition" , IEEE Transactions on Neural Systems and Rehabilitation Engineering Volume: 25, Issue: 9, pp: 1581-1591, Sept. 2017.

[6] D. Yu, K. Yao, H. Su, G. Li, and F. Seide, "KLdivergence regularized deep neural network adaptation for improved large vocabulary speech recognition," In Proc. ICASSP'13, pp. 7893– 7897, 2013

[7] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet ofThings (IoT): A vision, architectural elements, and future directions,"Future Generat. Comput. Syst., vol. 29, no. 7, pp. 1645–1660,Sep. 2013. [8] H.-W. Hon, "A survey of hardware architectures designed for speech recognition," Dept. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-91-169, Aug. 1991.

[9] L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993, pp. 1–9.

[10] D. A. Sunderland, R. A. Strauch, S. S. Warfield, H. T. Peterson, and C. R. Cole, "CMOS/SOS frequency synthesizer LSI circuit for spread spectrum communications," IEEE J. Solid-State Circuits, vol. 19, no. 4, pp. 497–506, Aug. 2019.