THINK INDIA JOURNAL

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

HUMAN COMPUTER INTRACTION USING SPEECH RECOGNITION TECHNOLOGY

Upinder Kaur
Computer Science Department
Lovely Professional University
Jalandhar, India
upinderkaur45@gmail.com

Nandini
Computer Science Department
Lovely Professional University
Jalandhar, India
Nandini.23493@lpu.co.in

Mehak
Computer Science Department
Lovely Professional University
Jalandhar, India
mehakkatnoria93@gmail.com

Swati
Computer Science Department
Lovely Professional University
Jalandhar, India
rampalswati@gmail.com

Abstract

Communication between a machine and man is provided by a technology called machine recognition technology. In this paper the speech recognition technology is discussed. This technology is implemented by comparing input string with the built-in dictionary. The security feature is provided here which asks random questions when the user logins. To make the user interact with the machine a database dictionary is created. Along with that input and output database is also created. Input database provides interfaces which provides user the input into the machine by speaking and then matching is performed with the inbuilt- dictionary. If the string is matched then it generates the output. In this, the machine recognizes the voice and generates output.

Keywords: speech recognition, speech-to-text, personal voice assistant

I. Introduction

The task of speech recognition technology is to recognize the words spoken by a user through a microphone and convert it into text form. It is also referred as speech-to-text [1]. Input is provided through voice. Various input forms are given by click-on mouse, keyboard typing. Similarly in speech recognition technology, input is provided through voice by talking on microphone. To perform speech recognition a software component is required which is known as engine for speech recognition [2]. The role of this engine is to process the voice input into text. Communication between people and machines is achieved through this speech technology. The communication should be such that the machines should respond in such a way as human being is responding. Machine should read facial expressions, understand body language and other

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

gestures done by humans. The concentration, emotion and interest should be same when communication is between human being and machines. Machines should also show the same intelligence level as a human being. In order for all this to happen, more dedicated research needs to be done in the speech technology field. The main task of speech recognition engine is discussed in this paper. The user login-ins and random questions are asked by the machine to the user. The engine then understands the voice input which in-turn is interpreted the result of recognition as command. The application is known as command and control application.

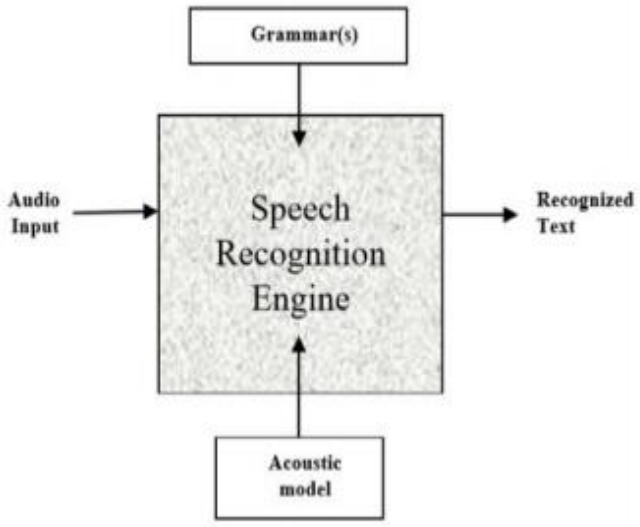


Figure: -Block diagram of working principle of Speech recognition engine

This paper is divided into four sections. Section I is on the introduction of human interaction with machine as artificial agents, Section II is on related research work, Section III is on proposed methodology about the interaction of humans with their speech vocabulary or voice recognition and Section IV is about the sum up of the proposed analysis of the work.

II. Literature Review:

Junaid Ahmed Ansari et al. [1] proposed a kit which is a voice command interface. The development of the kit has been done by integrating hardware components with open-source software. The kit contains package having replaceable components. Along with that, a C++ library is also provided which supports the application development. Also, the utility kit is demonstrated.

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

Leandro Yukio Mano et al. [2] proposed approach for identifying emotions in voice patterns. It becomes easier for the artificial agents to interact with machines, once the users emotions are known to them. However, providing computing systems with the capability to acknowledge and interpret the emotion of their users, is one of the great challenges in the area of Human-Computer Interaction. For the reason specified, the article adopts an approach based on the Ensemble of Classification which is concerned with identifying and classifying emotions on the user's motor expression basis. When this procedure is compared with the classical approaches adopted in the literature, the results show that, as well as achieving a high degree of accuracy, the proposed model maintains a good level of consistency when identifying the emotions of the users.

Efthimios Alepis et al. [3] discovered security and privacy towards voice assistants. To enrich the experience of the user, most repair supplier's square measure step by step shipping sensible machines with voice controlled intelligent personal assistants, to reach a new level. Although these systems make the device more intelligence but also there is high possibility of risk involved. Permissions models has been introduced, which informs user about the privacy resources which that application requires. This is how the security issues are countered and the risk is accessed.

Grace M. Begany et al. [4] discovered explored variations in the response of users to an oral communication search interface via voice input in comparison with a matter input search interface

D.G. Childers et al. [5] expressed interaction between Measuring and modeling vocal source-tract. Two factors which affect the synthetic speech are intelligibility and naturalness.

John H. L. Hansen et al. [6] has introduced the strategies which requires small amount of samples which will supply large amount of utterances. The first strategy which is used initial training to confine the estimated plan roughly. The second strategy has been used to create the samples as pseudo-whisper samples by using denoising autoencoders.

Vicente P. Minotto et al. [7] introduced a replacement approach for visual VAD and lips. Skin segmentation is the first step performed in the algorithm. It will cut back the search space for lip extraction. The technique is used to identify the lip motion named as Hidden Markov Models. A few parameter also has been captured using Audio information.

Arthur L. Robinson et al. [8] explored communication between system and human. A major motivation is to attain in man-machine interactions the potency of spoken language among humans. Continuous speech is tougher to know than are isolated words. Commercially

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

obtainable speech recognition systems of the latter kind are extremely productive despite their restricted capability. To acknowledge continuous speech, a lot of data is required than is contained in acoustic waves alone. The linguistic and discourse information that has got to be supplied or programmed into a system to accomplish speech interpretation is the subject of many analysis activities that are delineated. Speech synthesis systems face similar issues however are additional advanced.

Lu-Shih Alex Low et al. [9] described methods for Detection of Clinical Depression. In this study the various symptoms of depression were investigated with extremely huge number of 139 adolescents such as clinically depressed are 68 and controls are 71.

Subhasmita Sahoo et al. [10] discovered aggregation in voice. In this work, the key focus is on the second kind of aggression, i.e. Impulsive. This sort of aggression isn't reflected solely in words or phrases; it's conjointly exhibited by the affective state of the speaker that will are evoked by the interlocutor. However, during this research, only the affective state of the user has been though to observe aggression. Linguistic information was not taken under consideration.

III. Methodology

Proposed Work and Implementation:

→Personal Voice Assistant in Python

As we know Python is a suitable language for script writers and developers. Let’s write a script for Personal Voice Assistant using Python. The query for the assistant can be manipulated as per the user’s need. The implemented assistant can open up the application (if it’s installed in the system), search Google, Wikipedia and YouTube about the query, calculate any mathematical question, etc by just giving the voice command. We can process the data as per the need or can add the functionality, depends upon how we code things. We are using Google speech recognition API and google text to speech for voice input and output respectively. Also, for calculating mathematical expression Wolfram_Alpha API can be used. Play sound Package is used to play the saved mp3 sound from the system. Well, let’s get started with code. We will divide each function as a single code for easy understanding. Here’s the main Function, with get_audio() and assistant_speaks function. get_audio() function is created to get the audio from user using microphone, the phrase limit is set to 5 seconds (you can change it). Assistant speaks function is created to provide the output according to the processed data. So, we have got an idea here how we are giving voice to the machine and take input from user. The next step and the main step are how you want to process your input. This is just basic code, there is a lot of other algorithms (NLP) can be used to process the text in a proper manner. We have made it static.

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

Also, Wolfram_alpha API has been used to calculate the calculations part. Now we have processed the input, it's time for action. There are two functions included that is search_web and open_application. Search_web is just a web crawler which uses selenium package to process. It can search google, Wikipedia and can open YouTube. You just have to say include the name and it will open it in the Firefox browser. For other browsers, you need to install a proper browser package in selenium. Here we are using web driver for Firefox. Open_application is just a function uses os package to open the application present in the system.

Data Model: - Below diagram shows the engine of speech recognition internal processing.

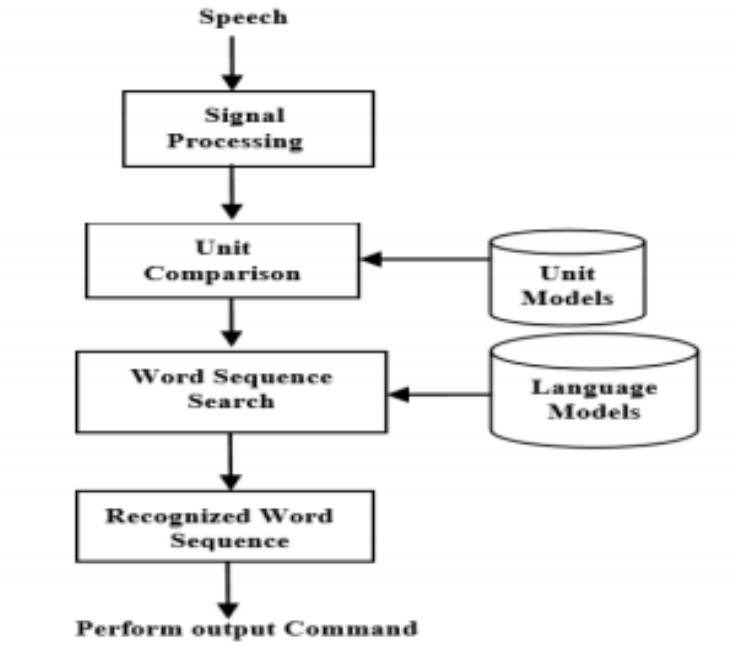


Figure: - The data modules of our speech recognition application

In the first step the speech is passed to the module which is used for signal processing. The speech is translated into patterns. It translates the speech waveform into speech patterns. The pattern further consists of series of feature vectors. This pattern is matched with the pattern stored with class identities. If the pattern is matched then the input voice is recognized and the command is executed and the output is generated. The output may be a response by speaking or an external interface. The processing is explained in the diagram shown on the next page.

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

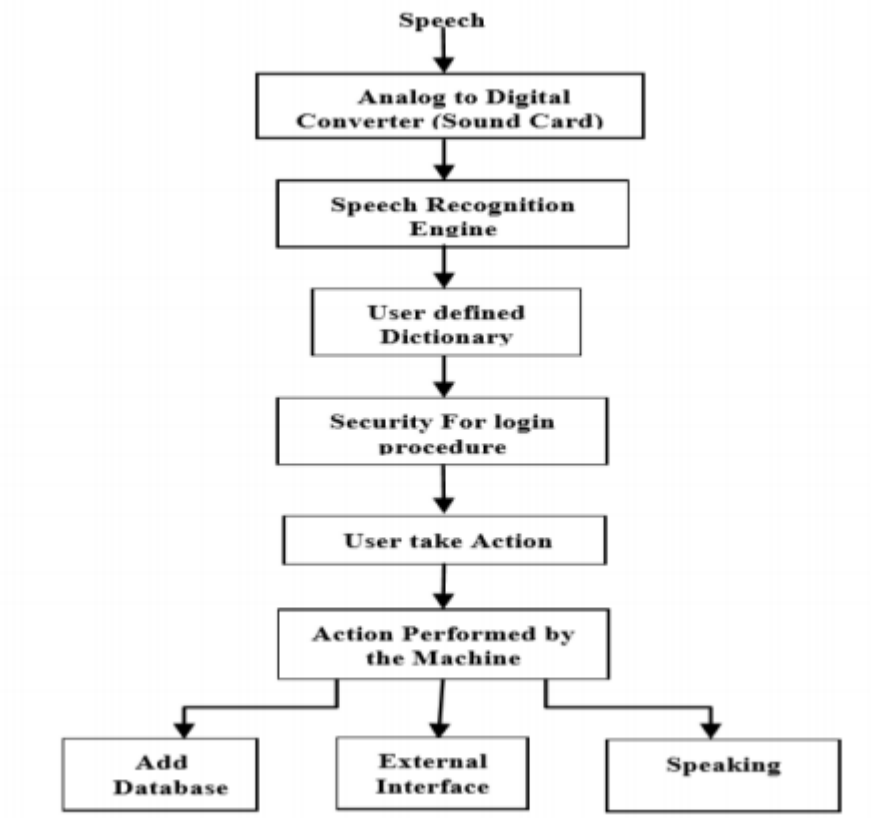


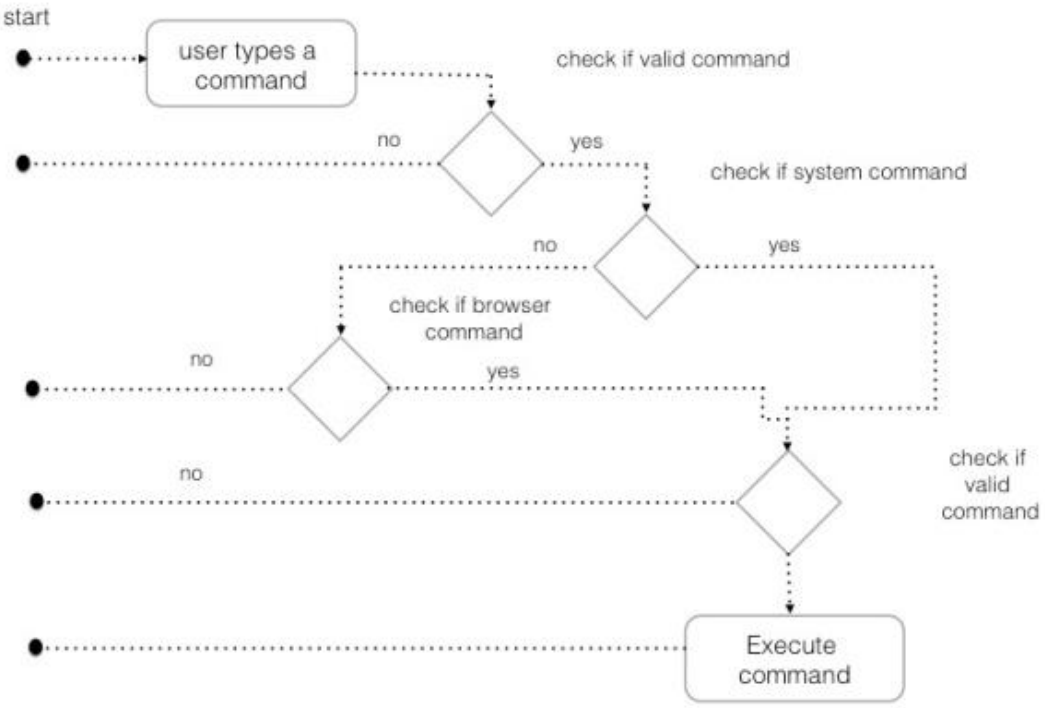
Figure: - The whole execution procedure of our speech recognition application

Through analog to digital converter, conversion of speech is done into digital signals. The application loaded with users voice is passed into speech recognition engine and converts it into digital signals. If the speech is then user can login on the system by answering random questions asked by the machine. The user then performs various tasks through speaking various commands. The response is given by machine based on the inputs given by user. There are various tools present like open notepad, open google and open command prompt. In addition to this, a huge recognition vocabulary is created for the user to interact with the machine. Along with this, a dictionary for the user is present in which all the useful commands are written and the interaction is done through commands.

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

Flow Chart: -



IV Conclusion

From this research works it is concluded that user can interact with machine through voice/speech. A human interaction application has been described using engine. Also, various commands are provided to user to make it easy for user to interact with machine. The user performs variety of tasks by speaking. In future, work could be done on challenges of human and machine interaction i.e detecting of emotions in voice patterns, understanding the body language and gestures of the person and responding back in the same way. These challenges are foreseen till now and more work could be performed removing these barriers in human and machine interaction. With the user experience to use the application will also increase.

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

REFERENCES

[1] Junaid Ahmed Ansari ; Arasi Sathyamurthy ; Ramesh Balasubramanyam: An Open Voice Command Interface Kit-IEEE Transactions on Human-Machine Systems- (Volume: 46 , Issue: 3 , June 2016)

[2] Leandro Yukio Mano ; Eduardo Vasconcelos ; Jo Ueyama: Identifying Emotions in Speech Patterns: Adopted Approach and Obtained Results - IEEE Latin America Transactions (Volume: 14 , Issue: 12 , Dec. 2016)

[3] Efthimios Alepis ; Constantinos Patsakis :Monkey Says, Monkey Does: Security and Privacy on Voice Assistants- IEEE Access (Volume: 5)

[4] Grace M. Begany ; Ning Sa ; Xiaojun Yuan: Factors Affecting User Perception of a Spoken Language vs. Textual Search Interface: A Content Analysis - Interacting with Computers (Volume: 28 , Issue: 2 , March 2016)

[5] D.G. Childers ; Chun-Fan Wong: Measuring and modeling vocal source-tract interaction-IEEE Transactions on Biomedical Engineering (Volume: 41 , Issue: 7 , July 1994)

[6] Shabnam Ghaffarzadegan ; Hynek Bořil ; John H. L. Hansen : Generative Modeling of PseudoWhisper for Robust Whispered Speech RecognitionIEEE/ACM Transactions on Audio, Speech, and Language Processing (Volume: 24 , Issue: 10 , Oct. 2016)

[7] Vicente P. Minotto ; Carlos B. O. Lopes ; Jacob Scharcanski ; Claudio R. Jung ; Bowon Lee : Audiovisual Voice Activity Detection Based on Microphone Arrays and Color Information-IEEE Journal of Selected Topics in Signal Processing (Volume: 7 , Issue: 1 , Feb. 2013)

[8] Arthur L. Robinson : Communicating with computers by voice- IEEE Transactions on Professional Communication (Volume: PC-22 , Issue: 3 , Sept. 1979)

[9] Lu-Shih Alex Low ; Namunu C. Maddage ; Margaret Lech ; Lisa B. Sheeber ; Nicholas B. Allen - Detection of Clinical Depression in Adolescents' Speech During Family Interactions-IEEE Transactions on Biomedical Engineering (Volume: 58 , Issue: 3 , March 2011)

ISSN: 0971-1260

Vol-22-Issue-17-September-2019

[10] Subhasmita Sahoo ; Aurobinda Routray : Detecting Aggression in Voice Using Inverse Filtered Speech Features - IEEE Transactions on Affective Computing (Volume: 9 , Issue: 2 , April-June 1 2018)