International Journal of Scientific Research and Engineering Development-– Volume 2 Issue 1, Mar-Apr 2019

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Voice Command System Using Raspberry Pi

M.H. Nandhinee[1] Akanksha khan[2], S.P.Audline Beena[3],Dr.D.Rajinigirinath[4]

[1,2] UG Students, [3]Asst professor, [4] Professor
Department of Computer Science and Engineering
Sri Muthukumaran Institute of Technology,Chennai-69,India

***Abstract*-Verbal correspondence is an imperative component in personal satisfaction, anyway upwards of 1.4 percent of people can't use common discourse dependably to convey their perspectives and sentiments with others which prompts discourse disadvantages. Discourse handicaps or discourse hindrances are the parts of correspondence issue in which the typical discourse get upset like faltering, lips, and so on. The word incapacity can turn away those individuals who are experiencing serious discourse inabilities from conveying in a method for doing thing that enables them to use for one closures their potential in instruction and entertainment. In this investigation another type of discourse acknowledgment framework is produced which perceives the disarranged discourse of the general population who are experiencing extreme discourse incapacities. Nowadays, users need a device or a communication system which is easy to handle. ASR systems work well for the people who are suffering from severe speech disabilities such as dysarthria which is the most common speech disorder and , these studies shows that there is inverse connection between level of weakness and exactness of speech recognition. This system describes the development of speech recognition system which recognizes the disordered speech.**

Keywords—Speech to text, Raspberry Pi, Voice Command System, Query Processing

I. INTRODUCTION

Talk is the best strategy for social correspondence. Only 5% to 10% of the human people has an absolutely customary strategy for verbal correspondence concerning distinctive talk features and sound voice; and whatever is left of the 90 % to 95 % experience the evil impacts of one issue or the other, for instance, wavering, scattering, dysarthria, apraxia of talk, etc. Wavering is a talk issue which is affecting a substantial number of people in their regular day to day existence. Wavering is an issue that interferes with plain talk.

A person who wavers may reiterate the underlying fragment of a word (as in wh what) or hold a singular sound for a long time as in (hhhheeeeelllllllooooo). Interventions, for instance, "um" or "like" can happen likewise; particularly when they contain repeated ("goodness ohh-ohhho") or deferred ("ohhhh"). More than 68 million people generally speaking stammer and has found to impact folks and females in the extent of 4:1. This issue is depicted by intrusions in the formation of talk sounds, called disfluencies. This technique makes it possible to use the speaker's voice and the spoken word to verify their identity and control access to services.

II. SYSTEM ARCHITECTURE

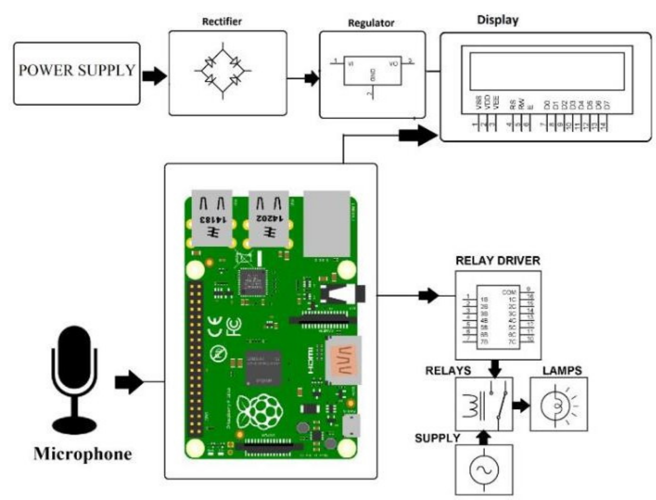
A. *EXISTING SYSTEM*

The current framework experiences the downside that just predefined voices are conceivable and it can store just constrained voices. Henceforth, the client can't get the full data reasonably.

B. *PROPOSED SYSTEM*

The proposed framework is to such an extent that it can conquer the downside of the current framework. The venture configuration include content to discourse. Here whatever the framework gets as contribution after the order the yield will get as voice implies discourse.

C. *HARDWARE IMPLEMENTATION*



Figure(1)

D. *MICROPHONE*

Amplifier is utilized to take the sound contribution of the sound. This sound information when further gone through the framework would be scanned for watchwords. These watchwords are fundamental for the working of the voice direction framework as our modules take a shot at the embodiment of hunting down catchphrases and giving yield by coordinating catchphrases.

E. KEYBOARD

Console goes about as an info interface fundamentally for the engineers, giving access to make alters to the program code.

F. MOUSE

Mouse additionally acts an interface between the framework and the engineer and does not have an immediate cooperation with the end client.

G. RASPBERRY PI

It is the core of the voice order framework as it is engaged with each progression of handling information to interfacing segments together. The Raspbian OS is mounted onto the SD card which is then stacked in the card space to give a working framework.

H. POWER

The Raspberry Pi needs a steady 5V, 1.2 mA control supply. This can either be given through an AC supply utilizing a small scale USB charger or through a power bank.

I. ETHERNET

Ethernet is being utilized to give web association with the voice direction framework. Since the framework depends on online content to discourse change, online inquiry handling and online discourse to content transformation consequently we need a consistent association .

J. SCREEN

Screen gives the designer an extra method to take a gander at the code and make any alters assuming any. It isn't required for any kind of correspondence with the end client.

K. SPEAKERS

When the question set forward by the client has been prepared, the content yield of that inquiry is changed over to discourse utilizing the online content to discourse converter. Presently this discourse which is the sound yield is sent to the client utilizing the speakers which are running on sound out.

III. STREAM OF EVENTS IN VOICE COMMAND SYSTEM

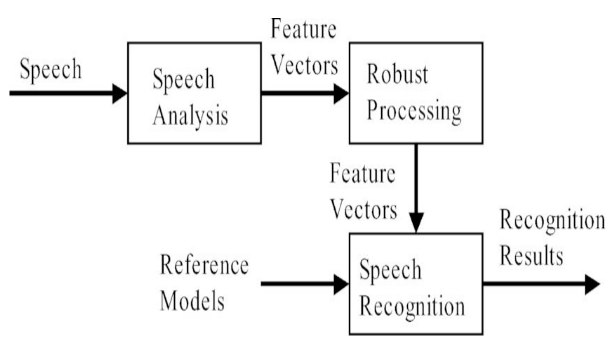
To start with, when the client begins the framework, he utilizes a mouthpiece to send in the info. Fundamentally, what it does is that it takes sound contribution from the client and it is sustained to the PC to process it further. At that point, that sound info whenever nourished to the discourse to content converter, which changes over sound contribution to content yield which is unmistakable by the PC and can likewise be prepared by it.

At that point that content is parsed and looked for watchwords. Our voice order framework is worked around the arrangement of catchphrases where it scans the content for watchwords to coordinate. Also, when watchwords are coordinated then it gives the significant yield.

This yield is as content. This is then changed over to discourse yield utilizing a content to discourse converter which includes utilizing an optical character acknowledgment framework. OCR sorts and distinguishes the content and afterward the content to discourse motor believers it to the sound yield. This yield is transmitted through the speakers which are associated with the sound jack of the raspberry Pi.

This module can be utilized to recover the subtleties of the status of the present framework under execution by utilizing the catchphrase "status". The data like current working framework adaptation, running CPU rate, number of CPU centers, framework name and current memory use is checked by utilizing the python framework and procedure utilities and the yield is given to the client in sound structure.

Flowchart of speech recognition:



Figure(2)

IV. MODULES IMPLEMENTED

A. DISCOURSE TO TEXT ENGINE

The Google discourse motor is a Speech-To-Text (STT) motor which is utilized to change over the directions given by the client in sound contribution to content structure, with the goal that these directions can be translated by the modules appropriately. To utilize the Google discourse motor, an application must be made in the Google designers comfort and the created API key must be utilized to get to the discourse motor. It requires ceaseless web association as information is sent over the Google servers.

B. CONTENT TO SPEECH ENGINE

CMU Flite (celebration light) is a little, quick run-time content to discourse amalgamation motor created at CMU and basically intended for little inserted machines as well as vast servers. Flite is planned as an elective content to discourse combination motor to Festival for voices fabricated utilizing the FestVoxsuite of voice building instruments. It is a disconnected motor and in this way web isn't required to change over the content reactions to sound reactions. It is one of the quickest motors accessible for use.

C. QUERY PROCESSOR

The Voice Command System has a module for question handling which works by and large like many inquiry processors do. That implies, taking the contribution from the clients, looking for applicable yields and afterward giving the client the fitting yield. In this framework we are utilizing the site wolfram alpha as the hotspot for executing inquiry preparing in the framework. The questions that can be passed to this module incorporate recovering data about acclaimed identities, basic scientific estimations, portrayal of any broad article and so on.

D. WIKIPEDIA

This module takes a shot at the catchphrase of "wiki". The framework requests what you might want to find out about. At that point the demand is made to the Wikipedia API for the required inquiry. It creates the synopsis of the data with respect to the inquiry and the information is yield through the amplifier to the audience in sound structure. If there should be an occurrence of disappointment, the blunder message is created saying "powerless to achieve lexicon of wiki".

E. OTHER COMMAND SPECIFIC MODULES

The Voice Command System also has some command specific modules like fetching hacker news, email and current time. Each of these modules is related to the system using keywords like “hacker news”, “email” and “time” respectively. Whenever any of this keyword is said to the system, it fetches that module and launches the contents of that module thereby providing the appropriate response to the user.

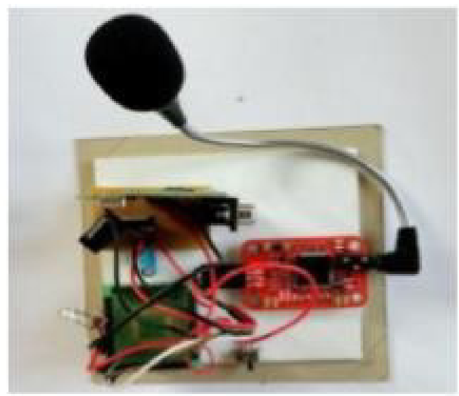
F. HMM

The idea was to recognize the input signal and compare results with manual annotations. Stuttering persons articulated earlier prepared sentences and the system recognizes them. Next, differences between automatic and manual transcription were found. The project design involves text to speech recognition. Here whatever the system receives as input after the command the output will get in the form of voice means speech. In order to perform tests the system using HMMs was build. It was based on context dependent models (triphones) with the adaptation to each recognized speaker.
Google API Technique is implemented to analyse the most probability of the word which is uttered by the person. Every attempt of pronouncing the word will be processed through the help of Google API technique and the uttered word is voiced out by the speaker. This reduces the stress of the person by avoiding the need to complete the sentence.

V. SIMULATION

A. Transmitter and Receiver sections

The product usage of different squares referenced in the past segment is finished utilizing the Embedded C. In the transmitter side, at first it will instate the microcontroller and the factors like port. On the off chance that the rationale 1 information is given to perceive, at that point it will send information 0x05 as push ahead direction to the beneficiary. So also, the other information will be given to perceive. In like manner, the transmitter will distinguish the direction in the product part and after that the comparing information will be sent to the recipient. At the recipient side, the microcontroller and the RF beneficiary are initialised to get the flag. On the off chance that the information got is 0x05, at that point the robot will perceive the order as push ahead and it will pursue the direction. So also, other information will be perceived.

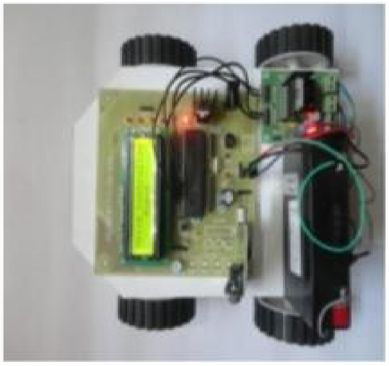


Transmitter section of stuttered speech recognition module

Figure(3)

B. Receiver section of stuttered speech recognition module

The recipient is the mechanized robot. It comprises of two DC engines and will make the robot move in forward or the retrogressive way. For the acknowledgment of faltered discourse, the test information is changed over to formats. The acknowledgment procedure at that point comprises of coordinating the approaching discourse with put away formats. The layout with the least separation measure from the information design is the perceived word



Receiver section of stuttered speech recognition module

Figure(4)

The best match (most minimal separation measure) depends on unique programming. This is called Dynamic Time Warping (DTW) word recognizer. Right off the bat, the voice order is put away in the Voice acknowledgment module with the assistance of the capacity keys accessible. This goes about as an example voice. At that point, the info voice is encouraged into the mouthpiece. It changes over the physical voice motion into electrical flag. The flag got is sustained into the voice acknowledgment module. The voice got is prepared in the voice acknowledgment framework and the attributes of this voice direction are coordinated with that of the pre-put away example voice. The voice acknowledgment module sends comparing string to the microcontroller, which is then transmitted through a RF module. The recipient gets the ideal voice motion from the RF transmitter through remote connection.

VI. RESULT

The Voice Command System takes a shot at the thought and the rationale it was planned with. Our framework utilizes the catchphrase "tune in" to take a direction. Every one of the directions given to it is coordinated with the names of the modules written in the program code. In the event that the name of the direction matches with any arrangement of watchwords, at that point those arrangement of activities are performed by the Voice Command System. The modules of Find my iPhone, Wikipedia and Movies depend on API calling. We have utilized open source content to discourse and discourse to content converters which give us the highlights of adaptability. In the event that the framework can't coordinate any of the said directions with the gave catchphrases to each order, at that point the framework apologizes for not ready to play out the said assignment. All things considered, the framework takes a shot at the normal lines with every one of the highlights that were at first proposed. Moreover, the framework additionally gives enough guarantee to the future as it is exceedingly adjustable and new modules can be included whenever without exasperating the working of current modules

VII. CONCLUSION

In this paper, we have successfully given the idea and support behind the Voice Command System. We have moreover cleared up the imperfections in the present system and our strategy for settling those deformities. Moreover, we have similarly spread out the structure designing of our Voice Command System. A significant part of our modules are of open source systems and we have revamped those modules as demonstrated by our structure. This gets the best execution from the system to the extent space time multifaceted design. The Voice Command System has a giant degree later on. Starting at now, we are seeing modest aides like Siri, Google Now and Cortana end up surely understood in the adaptable business. This gains the ground smooth to a complete voice request structure.

VIII. REFERENCES

[1] Juray Palfy, Jirt Pospichal, "Pattern Search in Disfluent Speech", 2012 IEEE International Workshop on Machine Learning for Signal Processing, Sept 23-26.2012, Santander, Spain.

[2] Ravi Kumar. K. M, Ganesan. S, "Comparison of Multidimensional MFCC Feature Vectors for Objective Assessment of Stuttered Disfluencies". Advance Networking and Applications, Volume: 02 Issue: 05, Pages: 854-860(2011).

[3] Cullinan. W.L, Prather. E.M & Williams. D, "Comparison of Procedures for Scaling Severity of Stuttering", Journal of Speech and Hearing Research, Pp. 187-194, 1963.

[4] Oliver Bloodstein, "A Handbook on Stuttering", 5th Edition, Singular Publishing Group, Inc., San-Diego and London, 1995.

[5] M N Hegde, Deborah Davis, Text book on "Clinical Methods and Practicum in Speech Language Pathology", 5th Edition, Cengage learning publisher, 2005.

[6] Andrzej Czyzewski, Andrzej Kaczmarek, Bozena Kostek, "Intelligent Processing of Stuttered Speech", Journal of Intelligent Information Systems archive, Volume 21, Issue 2, Pp. 143-171, September 2003.

[7] K. Ravikumar, B. Reddy, R. Rajagopal, and H. Nagara, "Automatic Detection of Syllable Repetition in Read Speech for Objective Assessment of Stuttered Disfluencies", Proceedings of World Academy Science, Engineering and Technology, Pp. 270-273, 2008.

[8] Hariharan.M, Vijean.V, Fook. C.Y, Yaacob. S. "Speech stuttering assessment using sample entropy and Least Square Support Vector Machine". Signal Processing and its Applications (CSPA) 2012, Pg no:240-245, ISBN :978-1-4673-0960-8.

[9] Marius Cristian, Adrian Graur, "Developing a Logopaedic Mobile Device Using FPGA", 1-4244-1234-X, 2007, IEEE.

[10] Walter Dosch, Nontasak Jacchum, "Stuttering Removal-Developing Mealy and Moore Style Implementations of an Interactive Component.