

Voice Based Intelligent Media Access System

U. Ravi Kumar¹, Rohit Patil², C. Pradeep³, R. Prakash⁴
School of electronics and communication engineering, VIT university, chennai.

Abstract -A voice command is given by human to control the media files. This voice command will eliminate the need of using buttons and switches by the consumers. There are many open source packages available for voice based speech recognition like cmu sphinx, htk, Julius etc. The outcome of the project is to develop an embedded system platform with linux os that is capable of recognizing speaker's commands in a specific domain. Currently we are accessing the media files present in the system, based on the voice command like songs.

We are using IOT platform in this voice based media access system to access a particular song that is not present in the system, as IOT applications are increasing rapidly in today's world.

Keywords—Speech recognition system, feature extraction, open source system

I. INTRODUCTION

Voice recognition is the process of access the speaker's word as a input to the Raspberry pi and this process provides a easy way of controlling the system. Voice command provides a way to control the system which allows the user to communicate with the system.

The most popular voice recognition software is the pocketsphinx. Speech will be recognized clearly if the microphone is placed within range to the user like for example if the user is wearing a microphone. Speech recognition eliminates the necessity to sit at a keyboard or work with a remote control. Voice recognition technology has many advantages. For example, a Person who ia an illeterate can use this application to complete their work like printing the pass book in the bank ,enquiring about the balance etc. .

The dis-advantages are as follows. To use speech recognition technology one has to raise their voice than their normal voice, in case there may be a vocal cord injury at that time the speech of the user cannot be recognized and an error occurs. The second dis-advantage is the person when he wakes upp from his sleep and tries to communicate with system, his voice may change if he would have shouted last night. This causes some in-convenience to the speech recognition technology.

If occasionally you have a trouble understanding what is said, it is obvious that a computer will have trouble much more often than us.

II. CONTROL FLOW.

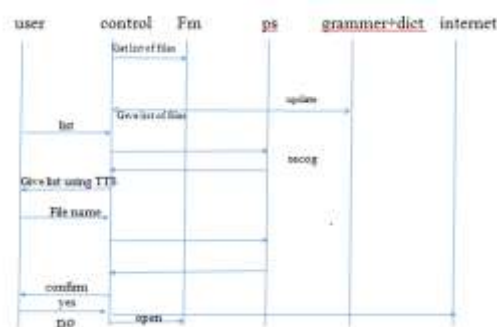


Fig1:Control flow diagram.

The fig:1 explains the flow of process fom one module to another module. User is giving a voice command to list the files, then the control will be transferred to the file manager. This file manager will update the grammar and the dictionary file with the help of festival software to create the dictionary file automatically. After updating the grammar and the dictionary file then the pocket sphinx will give the list of files present in the system. This list uses the text-to-speech to return the list to the user. If a particular file of a song is not present in the computer then the system will be redirected to the internet to search the song or a file in the internet.

The voice recognition is done on the raspberry pi using the pocket sphinx software .For recognizing we have to install the following

- Raspbian OS
- Sphnix base
- Pocket Sphinx

And after installing these software's into the PI we have interfaced the microphone.

III. TEXT TO SPEECH

The above mentioned system converts the voice given by the user to the speech which can be recognized by this software. This can be connected by small piece of data received and can be stored in the data.

TTS has two parts. First one is the front end and the other one is the rear end. The former does two tasks. The first one it does is that it transforms the text like numerical into words. The second one is the process of normalizing the numerical text. The later end gives the phonetic transcriptions to a specific word and breaks into small parts and transform into sentences.

IV. WORKING OF POCKET SPHINX

Voice recognition using pocket sphinx software. In this we have generated a dictionary file and to this dictionary we have generated a grammar file. We are able to recognize the voice given by the speaker.

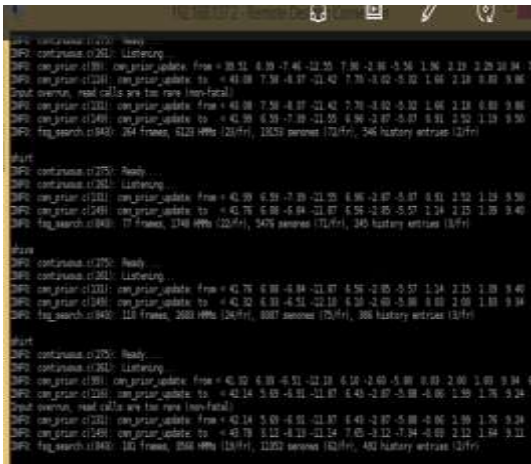


Fig2: Result using voice recognition.



The fig 2 shows the result we have obtained on giving the commands.

V. TESTING THE AUDIO CARD

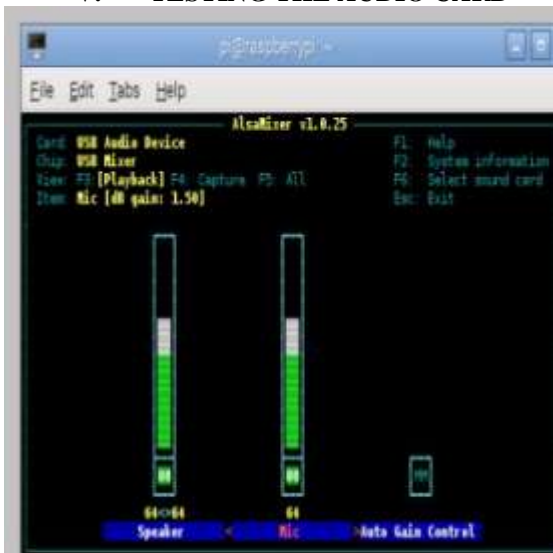


Fig:3Alsa mixer

To perform the speech recognition technology we have to check the working of of the microphone,if it doesn't we need to make it proper. After testing the mike by increasing the volume we need to list the connected usb ports to the raspberry pi.And che the microphone in that list.

We can adjust the volume of the microphone by pressing the up and down keys.The volume to be recorded can also be varied as per the arrow keys.After detecting the mike in the list make the microphone to detect the voice to the maximum it can.

VI. FESTIVAL SOFTWARE

Festival software is generally multi language software. This software works on the principle of text to speech system which supports many application program interface which recognizes the speech.Festival software supports many languages.

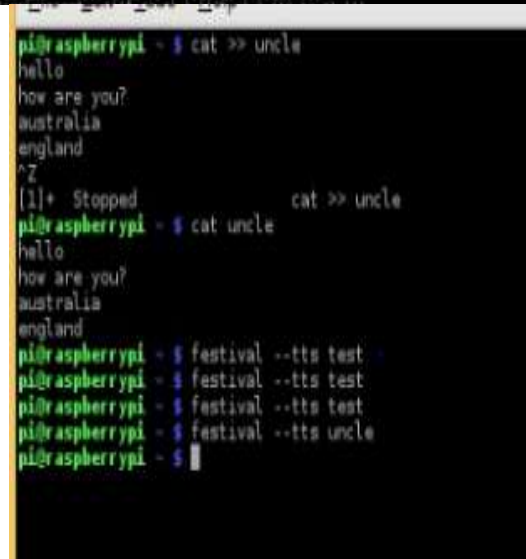


Fig:4 Output for festival software

The figure 4 explains the application of the festival software. In the figure we are giving some text to the file and then we are able to hear the voice in the form of TTS.

VII. COMPARISION BETWEEN FESTIVAL AND ESPEAK

Espeak is a text-to-speech or speech synthesizer software. There is a graphical user interface.

espeak -v en "Hello i am espeak"

This is the command line used for espeak software.

In this speech recognition technology we are comparing the festival and espeak software's for creating the automatic dictionaries. Working with espeak is a complex task so we are working with the festival software as it generates dictionary automatically.

VIII. INTERFACING VOICE WITH THE INTERNET.

The software we are using in this speech recognition system is google voice API. The voice commands from the user are given which are to be detected by the usb microphone. This voice command is converted using the speech to text software and converts the voice into the text. This text is then compared with the commands already written in the

nano file by the user. If the text matches with a particular command then that command will be executed. In this way we are interfacing the voice with the Google API. The block diagram for interfacing the voice with the Google is shown below:

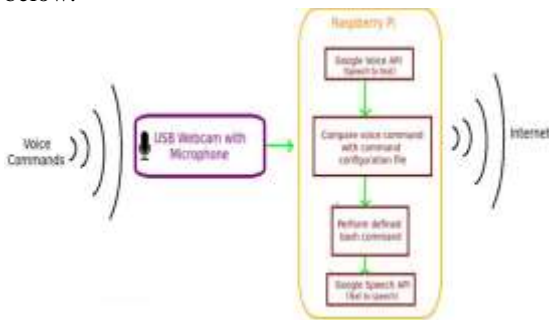


Fig5: Block Diagram.

IX. DICTIONARY GENERATION FROM INTERNET SITES

Using a google API for speech recognition as a drawback as since it requires seamless internet connectivity. we proposed to develop a grammar and dictionary database for the recent popular names. In this case internet is required only to get these names the remaining speech recognizing part only done on board using pocket sphinx and festival. festival software is used to generate pronunciation for the dictionary with respect to this songs of a particular artist have been taken and a dictionary for that songs and a grammar file has been created This dictionary will help in searching a song which is not present in the system .

X. PLAYING THE MEDIA FILE

For playing the media file present in the pen drive we need to write the shell script for the songs present in the USB drive. To run this shell script we need to install a voice recognition software that runs the shell script.



Here for executing the media we have generated a dictionary and a grammar file automatically. Using these two files and pocketsphinx command we have given the voice and played the media.

XI. REFERENCES

- [1] Kijima, Y.; Nara, Y.; Kobayashi, A.; Kimura, S. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84. pages 405-408
- [2] M Fazeen, G Bajwa, R Dantu – Context-aware multimedia encryption in mobile platforms, 2014 – dl.acm.org pages 53-56
- [3] <http://www.cmusphinx.sourceforge.net/html/cmusphinx.php>. Cole, R., Fanty, M., Muthusamy, Y., & Gopalakrishnan, M. Satori, H., Harti, M., & Chenfour, N.. Arabic Speech Recognition system based on CMUSphinx,2007
- [4] <http://wolfpaulus.com/journal/embedded/raspberrypi2-sr-2014>
- [5] <http://wolfpaulus.com/journal/android-journal/artist-voice-reorganization-2014>
- [6] <http://diyhacking.com/best-voice-recognition-software-for-raspberry-pi/>

