iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** IV **Month of publication:** April 2016

DOI:

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Embedded Voice Controlled Computer For Visually Impaired and Physically Disabled People Using Arm Processor

SankaraNarayanan.S[1], Vinothkumar.R[2], Prakash.A[3]
[1,2]B.Tech student, [3]Associate Professor, Department of IT, JCE, Chennai, Tamilnadu, India

Abstract: This paper gives the design and implementation of embedded voice controlled system using ARM cortex A7 for visually impaired and physically challenged people. ARM cortex A7 has features such as low power consuming than its predecessor and based on ARM v7 architecture, also it is a quad core processor. The system uses Google's speech API for the conversion of speech to text and text to speech which has the dictionary of phoneme for more than 60k words and many languages are supported.
Keywords: voice control system, speech recognition, embedded system using ARM processor, speech-to-text, text-to-speech.

I. INTRODUCTION

The 2011 census showed that 1.7% of the total Indian population were visually impaired people in especially urban areas, their problem is that there is no proper interface or method to access computer, therefore we have been building an voice recognition system for accessing an controlling a computer. There have many systems proposed throughout the discovery of voice recognition from the early 2000's for many applications and the voice control system using DSPIC control can handle 256 words and it uses Radio waves transmission so the SNR calculation may delay the output or the signal may even got last,Voice control systems with local speech engine requires lot of tedious works for Acoustic model Building,As every speech recognition clearly should not uses more resources locally and it should has faster response time.

Our paper tells the usage of google speech API for speech to text conversion and text to speech conversion which uses DNN-HMM for Acoustic model as it can handle any words up to the length of 100 characters and it also doesn't require user features in the spectrum of speech.

II. RELATED WORKS

A lot of research has been completed for visually impaired people. Some of the recent research methods are discussed here.

Suma Swamy and K.V Ramakrishnan proposed a paper for an Efficient Speech Recognition System [1], Their research uses the HMM model for acoustic model for providing efficient speech and speaker Identification and provides the base for many voice recognition applications and it uses the features such as MFCC from user speech with noise cancellation and followed by Vector Quantization (VQ) to code HMM Database Where lot of words can be recognized easily at efficient rate and it overloads the processor To detect simpler words with difficult phonemes and training of HMM requires a lot of speech corpus and impossible to collect that much of data.

S.Nareshkumar,N.Mariappan, K.Thirumoorthy proposed a paper for Database Interaction Using Automatic Speech Recognition[2] which uses Sphinx4 voice recognition engine for speech to text conversion for interaction with database where it requires lot of training to build acoustic model for efficient speech recognition and it also overloads the processor because every word is recognized locally and user cannot have an option to customize the system upon his/her needs and it has more error rates even upon vigorous training of words and each sentences although it only has language model for only 335 words and 1212 sentences.

Liang WANG, Jie ZHU, Yao LV proposed a paper on An Improved Method for Predicting Fundamental Frequency Contour in Mandarin Text-to-Speech System with a Small Corpus[3] Which uses corpus based clips for different sounds in the languages and the usage of two types of corpuses one based on Tonal frequency and High Frequency. The pitch jitter which is important and calculated based on GMM where the Negotiated pitch is adjusted by PSOLA algorithm which makes the sound more natural ,Thus GMM model may not provide an Average level of tts for other languages.

Tapas Kumar Patra, Biplab Patra, Puspanjali Mohapatra proposed a paper on Text to Speech Conversion with Phonematic Concatenation[4]. Thus the recorded phonemes in the temporary file directory are reconstructed based on the Words and the

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

accuracy is not certain and also impossible to collect that amount of phonemes.

III. PROPOSED SYSTEM

The proposed system mainly focuses on building an voice controlling system for physically challenged people to control the entire computer with and additionally for controlling robots. The components used in the project are Raspberry pi 2 model B (ARM processor with 4 usb ports, 40 GPIO pins for interfacing with External devices and 5v micro usb slot for charging.)with internet connection ,a microphone , sound card(general technologies, cmedia technologies).

The Architecture for the proposed system is presented in the following figure 4.1

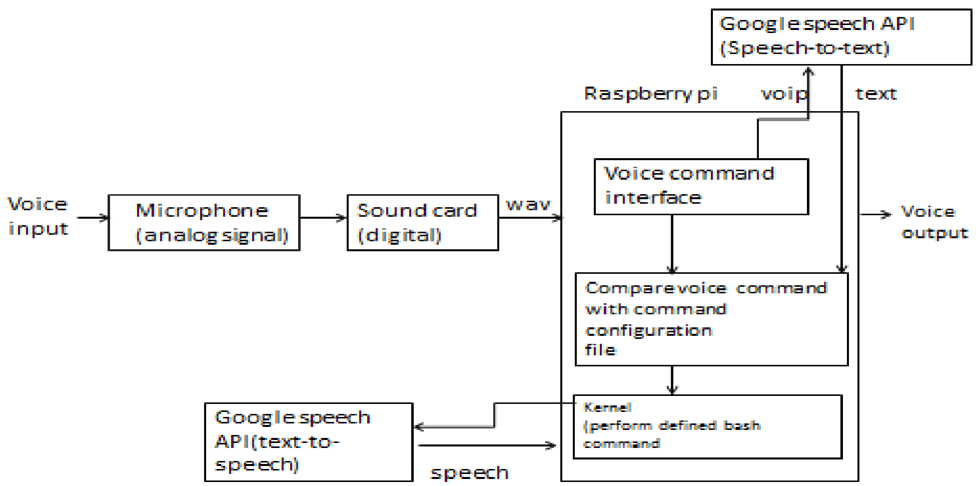


Fig 4.1 Architecture Diagram

The proposed system involves the following modules:

Speech Recognizer
Text-to-speech

A. Speech to Text

1) Features: Uses Google Voice recognition API so it has accuracy over 81% and Provides Large Vocabulary Continuous Speech recognition (LVCSR) and it is fast enough because of using Machine Learning Neural Networks algorithm for Acoustic modelling.

It has Quad core 1GB RAM ARM A7 microcontroller Raspberry pi with Raspbian Jessie os.

It requires less power than its predecessors and also releases less heat.

The user voice is recorded with microphone and the Sound card converts the user speech into WAV Audio file by using the help of raspbian jessie os on raspberry pi ARM7 microcontroller and then the converted file is stored in the temporary directory then our Application Voice command takes the WAV file as input with down sampling up to 32000Hz and the raspberry pi is connected to google voice API via internet which converts the speech to text and thus the API is effective in decoding because it uses Deep Neural Networks for Acoustic modelling instead of Regular HMM so it doesn't requires user features in speech as required with HMM and it process the speech and send the proper text command value to call a shell script to provide inputs to motors for their operation according to user commands.

The Commands we provided are

MUSIC

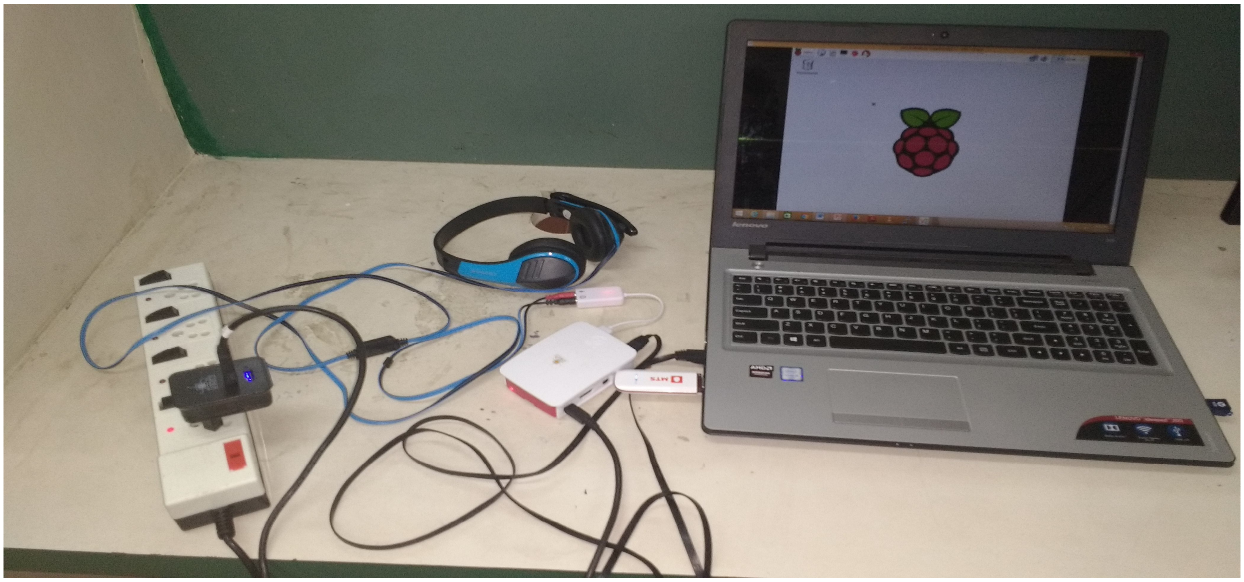
NOTEPAD

BROWSER

REBOOT

SHUTDOWN

The Figure 4.2 raspberry pi 2 model b



B. Text To Speech

1) Features: This module uses Google text to speech API.

In this the text will be converted to speech by the google

It powers applications to read aloud (speak) the **text** on the screen.

For this project we use the google text to speech module to provide notifications for the visually impaired people about the state of the system. Thus The Google Voice API for text to speech conversion does the job of reading the text from document on the screen for the visually impaired people.

IV. EXPERIMENTATION RESULTS

The speech recognizer circuit has an efficiency of detecting 85% of words for males and 93% of words for females. Thus we calculated efficiency for several words on 30 speakers (14 male, 16 female) and the experimentation results are as follows in Fig 5.1

Word	**Male**	**Female**
Read	89%	92%
Write	92%	94%
Open	95%	96%
Pi	70%	68%
Shutdown	81%	83%
Reboot	88%	90%
Music	85%	89%

A. Experimentation Inferences

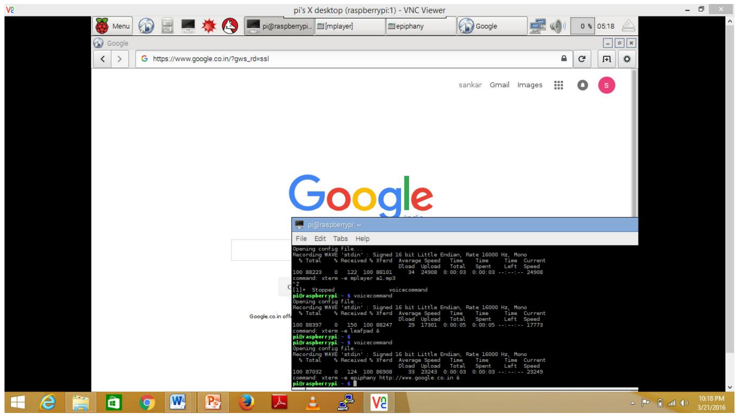


Fig 5.2 voice command to open browser

V. FUTURE WORKS

Features sending and receiving email through voice recognition and query processing in google can be the future enhancements.

VI. CONCLUSION

Our system may provide an efficient design of voice controlled computer and can be used with voice controlled Robots thus usage of Google API lightens the intensity of payload on processor and Text to Speech module useful in Gaining Basic Knowledge .

VII. ACKNOWLEDGMENT

We would like to thank Dr.A.Prakash who inspired us and guided us to involve in research work. We would also like to thank Dr.A.VijayaKumar for motivation and support. We are thankful to all those who helped us in creating voice control module.

REFERENCES

[1] Suma swamy and K.V.Ramakrishnan" AN EFFICIENT SPEECH RECOGNITION SYSTEM",Computer Science & Engineering: An International Journal (CSEIJ), Vol. 3, No. 4, August 2013.
[2] S.Nareshkumar, N.Mariappan, K.Thirumoorthy, “Database Interaction Using Automatic Speech Recognition”, International Conference on Innovations in Engineering and Technology March 2014
[3] Liang WANG, Jie ZHU, Yao LV proposed a paper on An Improved Method for Predicting Fundamental Frequency Contour in Mandarin Text-to-Speech System with a Small Corpus
[4] Text to Speech Conversion with Phonematic Concatenation",Tapas Kumar PatraBiplabPatra,International Journal of Electronics Communication and Computer Technology (IJECCT)Volume 2 Issue 5 (September2012)
[5] Ronald M. Baecker, “Readings in human-computer interaction: toward the year 2000”, 1995.
[6] http://jasperproject.github.io
[7] M.A.Anusuya, S.K.Katti,” Speech Recognition by Machine: A Review”, international Journal of Computer Science and Information Security,vol 6,issue 3,pp-181-205,2009