Real Time Speech Recognition and Home Automation Using FOSS* Tools

Arfat Bin Kileb[1], Viswajit Agam[2], Harsh Zavery[3], Tushar Sandbhor[4]
[1, 2, 3, 4] Department of Computer Science & Engineering
[1, 2, 3, 4] SRCOE University of Pune-412207

Abstract- *Speech is the first important primary need, and the most convenient means of communication between people. Also speech is core part of any intelligent system. The communication among human computer interaction is called human computer interface (HCI). This project basically focus on major technological perspective and appreciation of the fundamental progress of speech to text conversion and also gives complete set of speech to text conversion based on Raspberry Pi board and CMU Sphinx framework. A comparative study of different technique is done as per stages.*

This system is able to make a user's interaction with home equipment easier by providing an easy to use, interface. This system is capable of interpreting speech as commands. These commands can be formulated in several different ways to make the interaction more natural. After a command has been interpreted an appropriate action is taken on the correct home equipment.

Keywords- Embedded system, Home automation, Natural language Processing, Raspberry pi, Speech recognition, CMU-Sphinx.

I. INTRODUCTION

Speech recognition, often referred to as Automatic Speech Recognition (ASR), is a technology in which human speech is processed into text. Most people today probably associate speech recognition with computers. However, the first attempts to process speech signals were in fact done mechanically.

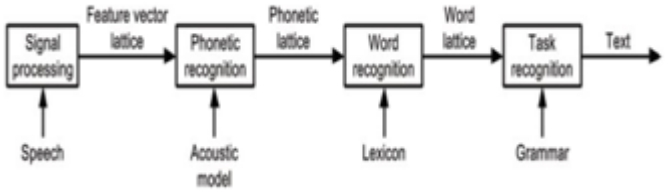

Figure 1: Speech recognition Architecture

With the advancements in Information Technology, the next generation of user interface is desired to be more user-friendly and powerful. As the choice for natural and expressive means of communication, speech is more desirable for the human-computer interaction. Furthermore, considering desktop PCs to mobile phones, toys and other embedded devices, the user interface becomes smaller in size which limits its operation. Speech has the potential to provide a direct and flexible interaction for the embedded system operations.

The input of the system characterized by human voice signal, detected by a USB microphone connected to the system. After passing through sequence of steps word gets recognized. The recognized word is converted to text and we will get the output in terminal window. Using GPIO pins of Raspberry Pi these texts are transmitted using a wired modem. One more wired modem at the receiver end will receive those words and according to that automation can be done

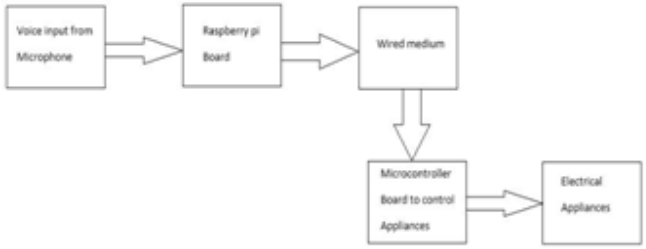

Figure 2: Home Automation

III. WRITE DOWN YOUR STUDIES AND FINDINGS

a. Outcome of study on speech recogni-tion software

The study was conducted on the leading open-source software for speech recognition: HTKspeech-recognition toolkit (HTK), CMU Sphinx. It was discovered that all software had little active development on their code base. This discovery made the existence of an active community an important factor since it shows that the software is still relevant. In order to motivate the choice of software all aspects presented in subsection 1.4.2 will be discussed separately and then combined for the final decision.

1. Programming language

The studied speech-recognition software is implemented in python programming language. while Sphinx

supports both C and Java with two different frameworks, PocketSphinx and Sphinx-4. Since all members in the project group were already well versed in Java, Sphinx-4 was the preferred software in this aspect.

2. Availability of models

When performing speech recognition some models are needed in order to interpret the speech input. The system needs an acoustic model and either a statistical language model or a grammar. In order to create acoustic models with high quality a large amount of speech data is needed.

Since not enough resources for creating new models were available during the project the existence of working models was very important. Most of the speech-recognition software aimed towards research focuses on enabling the use of multiple types of models rather than providing existing ones. HTK is mainly focused on research of hidden markov models in speech recognition is focused on research of acoustic models. Since most users create their own models during their research there is limited need to provide any.

Sphinx on the other hand provides multiple highly-trained models. Guides for further adaptation and training of these models are also available. Julius has satisfying models, though the acoustic models are not in English which made the usefulness of this software rather limited for this project. There exist some models that can be used with most speech-recognition software but since Sphinx has software-specific models it was preferred over the others.

3. Final outcome

After analysing all aspects, we decided to use Sphinx-4. Below follows a short introduction to how the framework is implemented.

Sphinx-4 consists of three main modules:

1. the front end,
2. the decoder
3. and the knowledge base, as seen in figure 3.

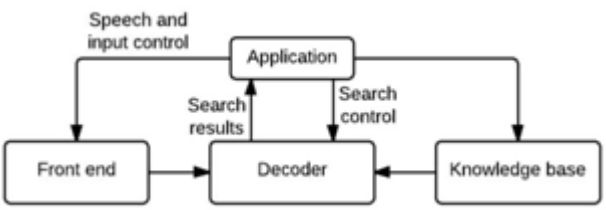

Figure 3: Sphinx-4 Architecture

Each of these parts are built with modularity in mind, so that the modules can be replaced depending on the needs of the developer.

VI. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the Importance of the work or suggest applications and extensions. The prototype developed during the project performed with satisfying results. The accuracy reached with users that has distinct pronunciation resulted in a limited need to repeat commands.

ACKNOWLEDGMENT

The authors would like to acknowledge the contribution to this work of all staff: Prof. Chandrashekhar Adki(Assistant Professor), Prof. Dipti Varshney(H.O.D.),Prof. Babaso Shinde(Assistant Professor) and all friends and would like to thank all of them for their support.

REFERENCES

[1] A.G. Veeravalli, W. D. Pan, R. Adhami, and P. G. Cox, A tutorial on using hidden markov models for phoneme recognition, in Proc. IEEE SSST05, Mar. 2005, pp. 154-157.

[2] CMU Sphinx Wiki, [online]. Available: http://cmusphinx.sourceforge.net/.

[3] GIT, [online]. Available: http://git-scm.com/.

[4] Gu, Yietal. ”Design and Implementation of UPnP-Based Surveillance Camera System for Home Security.” Information Science and Application (ICISA), 2013International Conference on. IEEE, 2013.

[5] Home Automation and Smart Home Control, [online]. Available: http://www.control4.com/

[6] IEE Signal Processing magzine entitled Large Vocabulary Continues Speech Recognition System. http://ieeexplore.iee.org/.

[7] IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 21, NO. 6, JUNE 2013 entitled Cross-Lingual Language Modeling for Low-Resource Speech Recognition: http://ieeexplore.iee.org/.

[8] International Journal of Emerging Technology in Computer Science and Electronics (IJETCSE)ISSN 0976-1353 Volume 13 Issue 2 MARCH 2015 entitled Modelling Of Secured Voice Recognition Based Automic Control System.

[9] J. G. Wilpon, Automatic recognition of keywords in unconstrained speech using hidden markov model, IEEE Transactions on Acoustic, Speech and Signal Processing, vol. 38, no. 11, pp. 1870-1878, Nov. 1990.

[10] Raspbian,[online].Available:http://www.raspbian.org/.

[11] Sphinx-4 Application Programmers Guide - CMUSphinx Wiki, [online]. Available: http:cmusphinx.sourceforge.net/sphinx4/.

[12] YanboZhao ; ZhaohuiYe A low cost GSM/GPRS based wireless home security system IEEE TransactionsonConsumerElectonics, (Volume: 5,issue:2)

***FOSS**: Free And Open Source Software (In case of this research paper we are using Sphinx-4, Raspberry pi as FOSS)