International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 2, March - April 2017 ISSN 2278-6856

Voice Controlling Linux Systems

[1]Brijesh K R, [1]Mahantesh S M, [1]Abhishek Roy, [1]Debanga P Bora, [2]Ms Suriya Refai Begum

[1]New Horizon College of Engineering
Department of Computer Science Engineering, Bangalore, Karnataka, India

[2]Senior Assistant Professor, New Horizon College of Engineering
Department of Computer Science Engineering, Bangalore, Karnataka, India

Abstract - ***This paper aims at implementing a voice based user interface for the Linux Desktop Operating System. Although Linux is predominant in the Server market, users of the Linux Desktop are less in number. The present methods of interacting with the Linux Desktop is through manual user input such as through keyboard, mouse, or remote network login using SSH or Telnet. The primary aim of this paper is to provide a new interface between the user and the Linux system by the means of user's voice.***

Keywords –Speech recognition, Pocketsphinx, Linux Desktop, Scripting

I. INTRODUCTION

With the advancement of technology, the dependence on machines for day-to-day activities has increased to such an extent today that we literally cannot live without them. One of the major necessities of any technology is its User Interface and ease of use. Today most of our devices such as smartphones, laptops, cars and even household appliances such as microwave ovens, refrigerators, etc. come with touch capabilities.

With the recent advancement in technologies, such a voice based user interface to machines is in the grasp of reality. Examples of such technology can be found in the form of Cortana from Microsoft Corporation, Siri from Apple Inc, Google Assistant from Google Inc, and Amazon Alexa from Amazon. These are all proprietary software/hardware products. But no such software exists for the Linux platform.

We decided to put this thought to effect. Linux is a strongly Command Line oriented Operating System and the Linux Desktop editions distributions or distros for short having many Desktop Environments or DEs. There are Desktop Enironments such as KDE a sleek looking DE to Gnome 3 and MATE, a fork and continuation of the now discontinued Gnome 2. Even with the recent accelerated development of these DEs, Linux Desktop has a small user base. But even with so many advancements in the GUI, Linux Desktop still uses the terminal for essential actions such as system administration, package management, etc. We decided to bring a new interface to the Linux Desktop by trying to automate these actions to some extent and enabling the user to communicate with the system in a new interface.

II. PROPOSED IDEA

As mentioned earlier, we are implementing voice control in Linux Desktop Systems. The proposed system consists of

1) The user's Voice as an input parameter.
2) A speech recognition engine which recognises and interprets the user's voice input and parses it into meaningful phrases.
3) If any command or action is present in the recognized phrase, then the system takes appropriate actions based on programmed instructions.
4) The system then notifies the user the outcome of the action executed in an visual and audible format using Text-To-Speech engines and appropriate notification messages.

III. RESEARCH AREA

The Carnegie Mellon University has developed an Open Source Speech Recognition Toolkit named CMUSphinx Pocketsphinx(1) which is be used to handle the speech recognition.

We found studies conducted on Voiced Based Login Authentication For The Linux System(2), but such an undertaking of implementing a new interface has not been done.

We decided to use Python 3(4) language and Shell scripts for coding the project. Since Python is an integral part of the Linux Desktop and since most of the Linux Distros utilize Python(4) for various purposes from GUI to many native applications being programmed in it. Linux system commands and calls are best executed using Shell Scripting. Python 3(4) and Shell Scripting are the scripting language used to code the instructions to the system since it is easier to call Linux system calls from both Python and Shell scripts as compared to other programming languages.

For GUI of the project, we decided to implement it using PyQt5. It is a Python wrapper for the Qt Framework. PyQt5 is a dual licensed software tool. It is both commercially licensed as well as Open Source licensed. PyQt5 renders a cleaner and much more appeasing front end as compared to other GUI frameworks such as Tk, wxpython, etc.

eSpeak is an open source software text-to-speech synthesizer that is available in many languages. In synthesizing speech, eSpeak uses Formant synthesis

method. This allows many languages to be provided in a small size. The produced speech is clear and can be used at high speeds, but it is not as natural or smooth as human speech.

IV. IMPLEMENTED TECHNIQUE

The architecture has three main components namely,

(1) The Speech Recognition Engine for user voice recognition,

(2) The programmed scripts with action instructions to be executed

(3) A Text-To-Speech engine for audio output.

The Speech recognition engine used is CMUSphinx Pocketsphinx. It is an Open Source Speech Recognition Toolkit developed by the Carnegie Mellon University(1). Pocketsphinx is imported as a module into the Python 3 scripting language and the user's voice is parsed from the Python script. The peripheral used for accepting the user's voice into the system is a microphone.

Python 3 and Shell Scripting are the scripting language used to code the instructions to the system. The actions to be taken when a particular phrase is recognized from the user's speech is to be coded using these two scripting languages.

eSpeak is an open source software text-to-speech synthesizer that is available in many languages. In synthesizing speech, eSpeak uses Formant synthesis method. This allows many languages to be provided in a small size. The produced speech is clear and can be used at high speeds, but it is not as natural or smooth as human speech.

The project will be an Open Source project with the final product being released for the general public under an Open Source license.

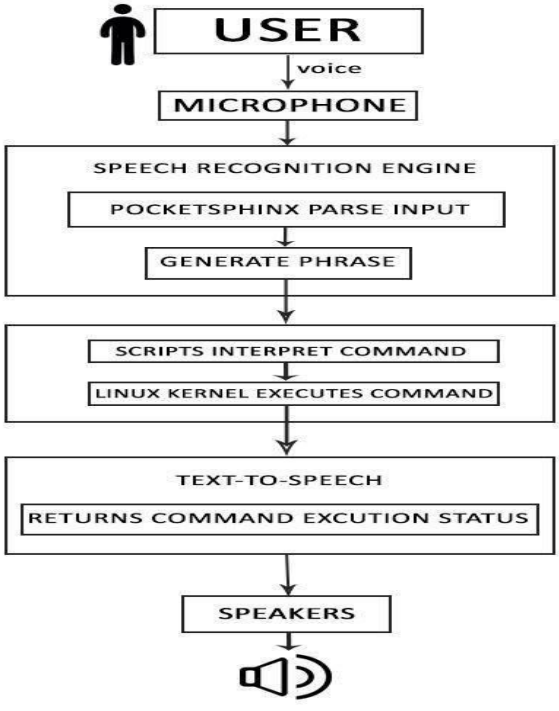

Figure 1 shows a basic flow structure of the project.

Due to complexity of certain Linux system commands, the use of keyboard cannot be avoided in full measure. But the use of keyboard will only be required in case of erroneous interpretation of the user's voice or when a complex input such as a package name which is a combination of alphanumerical with some special characters and regular expressions are to be entered as input.

Pocketsphinx does not need an active Internet connection for it to function, therefore providing offline usability to the proposed system. This makes the system effective for users without an Internet connectivity. The system will require a minimum of 4GB RAM, any latest Linux Desktop Distribution, 500-750MB of storage space for storing the various system libraries and dependencies that the proposed system will require. Due to various differences in the accent, dialect of users worldwide, in very rare cases, the user will have to train the built in acoustic model of Pocketsphinx to be able to recognize the user's voice. This may be applicable to users who have a strong influence of other non-English language in their English accent or different dialects of English.

V. FUTURE DIRECTIONS AND IDEAS

The project has a limited implementation possibility of a Virtual assistant, which can be implemented in future. The other major possibility is complete control of the system through the user's voice. There is also prospect of upgrading the voice output to a natural one.

Voice control and programming language could lead to future IDEs, which could revolutionize the Linux environment. Online Search & Response, Automated learning of System are a few possibilities if this project is taken further.

VI. CONCLUSIONS

The project provides a simplistic approach towards voice based system control in Linux systems. The proposed Voice control system gives an efficiency of about 80%. For speech recognition, the system is tested for words that are present in the dictionary. However the system will be more efficient in a generally quiet environment. The aim of this project in general is to provide the users of Linux Desktops a whole new interface to interact with their Linux systems.

REFERENCES

[1] CMUSphinx Pocketsphinx http://cmusphinx.sourceforge.net/https://github.com/cmusphinx/pocketsphinx
[2] IEEE Paper "Voice *Based Login Authentication for Linux"* by Sarabjeet Singh.
[3] Book "Building a Virtual Assistant for Raspberry Pi" by Tanay Pant
[4] The Python Foundation https://www.python.org/
[5] The Linux Foundation https://www.linux.com/https://www.kernel.org/
[6] PyPi https://pypi.python.org/pypi
[7] Espeak http://espeak.sourceforge.net/