

Automated System Using Speech Recognition

Siddhesh Sangale, Anant Bhagat, Tushar Patil, Prof. P. Y. Itankar

^{1,2,3,4}Department of Computer Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, Maharashtra, India

ABSTRACT

In today's world where technology is advancing to its highest degree, new updates are coming daily. Tasks that seemed impossible at first are appearing like a piece of cake today. As mentioned above, there are still some parts of tech that have to be taken a step ahead of its time. People still use mouse and keyboard for software access and modifications. As Computers can be operated with only keyboard and mouse, a new era of voice assistant has invaded the house of Technologies. In mobile we have Siri, Alexa, Google Assistant and in Computers we have Cortana as well. However, the computer technology needs to be updated to perform the various computer operations rather than just being limited to web search. So we have concentrated ourselves in making a Desktop Assistant that will perform the same task as that of siri, alexa i.e. web search along with the ability to manage and modify the files present in the system. In this project, we are making an automated system in such a form that can be provided directly to the system in executable format and will perform operations as soon as the project is activated.

Keywords: Automated System, Automated Desktop, Speech Recognition, Assistant

1. INTRODUCTION

Technology has changed the whole process of work in the past few years. From Computers to Mobiles to Smart watches, everything is evolving. However, People using systems like laptops or computers are still stuck on the use of keyboard and mouse for performing various tasks. We want to change this mode of connection to speech and control the whole system using the speech commands. We are focused on making a system such that not only the web part but also the files system of a particular machine must be accessed, modified and can be worked upon using speech commands. We don't want to restrict it with the kind of functions that are mapped for performing only web applications. We are focused on system that can perform file operations too.

1.1 Motivation

Automated systems can be very much easy to use for the users and are also fun and convenient. Such Automated systems or assistants are also taking its place in the professional domain such as businesses.

Disable persons face great difficulties while using computers due to physical limits. It is very unfortunate that people in this age also have to limit their abilities to interact with technology. Digital assistants help a lot of disable persons to overcome their disability to interact with technology.

All mentioned above caused us to think on this topic and work on it. As we know desktops and laptops can be accessed through keyboards and mouse. So this Idea struck of building a system that can perform more of Desktop operations and secondly web search operations and can further be researched and developed

1.2 Previous work

Being able to visually personalize the idea that now-a-days everybody has Google assistant, Siri and Alexa working for them however no one has a personalize assistant that can operate Laptop/Desktop for them except for Siri (Mac) and Cortana (Windows) which again is limited to performing the functions related to web search only. Assistance based mainly on certain types of web page search results but are deprived of working with the files in the machines. They have Limited functionality mainly of fetching information from the internet to derive results.

1.3 Application

Automated Desktop Using speech recognition is Computer program which helps users to communicate and make desktop perform/interpret operations as per the need. This System can help many physically disabled people to get in touch with Desktops / computers.

This one platform can change many people's user experience as they will be able to communicate with computers in an easier and convenient way. It will increase efficiency of the work by allowing them to perform multitasking.

2. LITERATURE SURVEY

From last few decades Speech Recognition has been through some major Innovations. Operations such as web-search, Dictation and voice commands have been Basic features on Smart Phones and Compact Devices.

2.1 History of Speech Recognition

If we took a glance in history, the speech recognition based device was first created by IBM in 1962 and was called as shoebox because it was size of shoebox. In 1970's Another Research on the Project was ongoing at Carnegie Mellon University in Pittsburgh, Pennsylvania with support of U.S. Department of Defense Advanced Research Projects Agency (DARPA) and came out with Harpy Which could Understand 1000 Words about a vocabulary of three-year old child. After few years in 90s that is in 1993, Some of the Big Organizations started Deep research on voice acknowledgement some of them are Apple, Google, IBM. Macintosh also started with Speech recognition project with its Macintosh's PC.

2.2 Current Status of voice assistants

At the Core of Speech acknowledgement, it is a synchronous Cycle of Voice commands and hear responses. Recently Sutar Shekhar and many more researchers came up with a application which has implementation of sending message through voice acknowledgement which Could help Visually Impaired. Omyonga Kevin and Kasamani Bernard Shibwabo have came up with an application which implements spoken commands even without internet connection providing flexibility over data costs. For development in Mobile technologies, Tong Lai Yu and Santhrusna Gande Have Implemented a ideology of Speech recognition in terms of Open Source services which can help many physically challenged programmers in development.

2.3 Aim of Project

On analyzing these systems, we came out with a conclusion, they are basically designed to work on Android and Desktop platforms but Only for Web-search Operations. So we Took an Initiative to Develop voice Recognition System For Desktop platform which could perform both Web-search as well as File implementation Operations On Desktop Platform. We have used python programming language which is a one of the very Robust Programming languages out there available and with help of Python Libraries Such as pyttsx3, nlptk(natural language processing library) and many more various libraries are used to make the development easier and work with good accuracy.

3. REQUIREMENT ANALYSIS

A. Hardware Requirement (minimum)

1. Processor – Intel Core i3
2. RAM – 4GB
3. 512 GB SSD / 1 TB HDD
4. Sound card with inputs for mic/headphone

B. Operating System

1. Windows 10 or above.

4. METHODS AND WORKING

4.1 Methods Used for making the system

1. Jupyter notebook: We have used Jupyter Notebook as an environment to write and test code. It is an open source application that gives you full access to write code, view and share documents.
2. Sqlite3: Similar to other databases, it comes integrated with python (2.5 +). This module is used to access the SQL database with a user-friendly interface and easy to remember commands. This is done with the help of API.
3. pyttsx3: This library helps to convert text to speech in python. Main benefit of this library is that it does not require an internet connection to convert text to speech.
4. Speech Recognition: This library helps us to convert speech to text in python. This library is responsible for speech recognition, breaking it into useful commands and displaying it to you. However, it requires a good internet connection to perform speech to text as it works upon an API.

4.2 Working of the system

Working of the system will follow up the specified instruction rules so has to complete the task

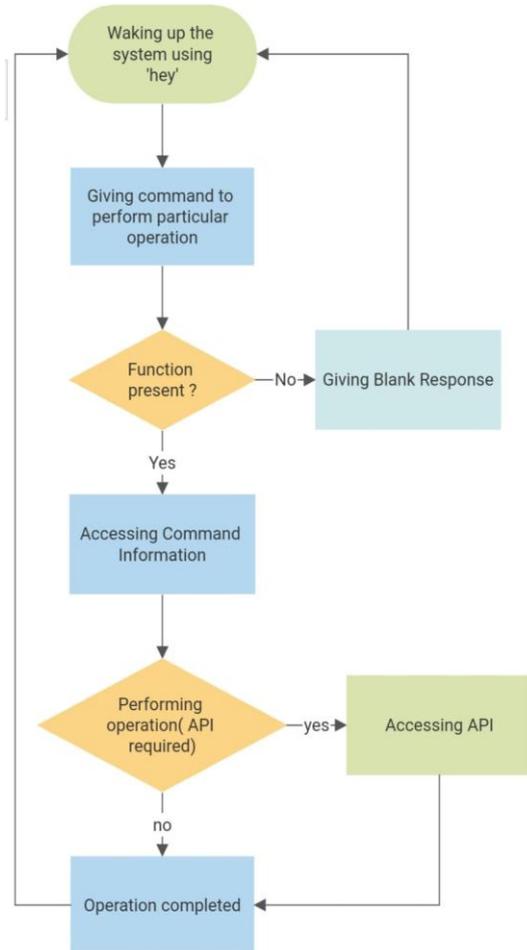


Figure 1. Working Flowchart

1. Waking Up System

To use the system we first need to call it using the Hey command. This is done to avoid the system taking surrounding noise as input and only to run it when we actually require it to perform the operations.

2. Command

After activating our assistant we will give the command related to the operation that we want to perform.

3. Function Present

This will check if the given input is right and is present in the operations directory. If 'yes' it will start to perform the commanded operation. If 'no' it will give the null response and will go to listening mode till it is again called.

4. Performing Operation

It will first access the function and will ask the user for more information if required. Then it will directly proceed to checking the needs required.

If API is being used in the commanded function, then it will request the information and will then fetch from API and will complete the commanded operation. Else if API is not required it will directly move to complete the task.

5. Output

After completing, output will be given with respect to the predefined method in function. It will then go to listening mode till it is again called.

Presenting Use Case Diagram of how overall system will be working.

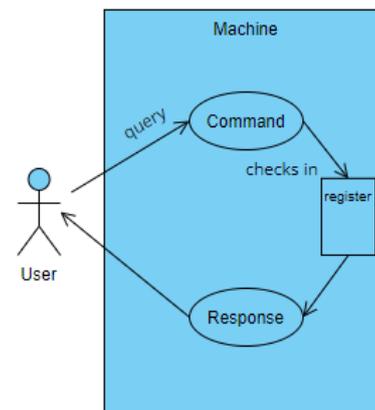


Figure 2. Use Case Diagram of Automated System

5. RESULTS AND DISCUSSION

As stated that most of the previous work focused on only web search portion of the system and completely neglected the files and windows automation we have focused mainly on the building and deployment of file search and file operation using voice commands along with leveling up the web search operations.

5.1 File Operations Implementation

1. File search and handling

First you have to select the drive or directory from where you are going to load the file or like where the file is present.

After Selection of drive this lets you to search the file using voice command and gives you in return all the files with the name you searched for. After it ask weather you want to perform task on single or multiple files or all files.

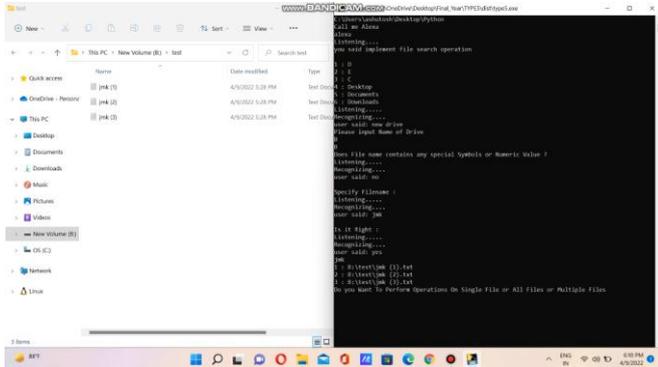


Figure 3. Searching File from specified drive

After selecting files it will lead you to choose one of the following operation : open, copy, move, delete.

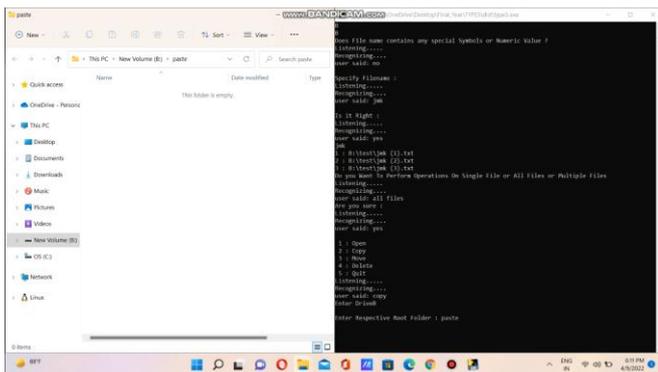


Figure 4. Copy Operation initialization

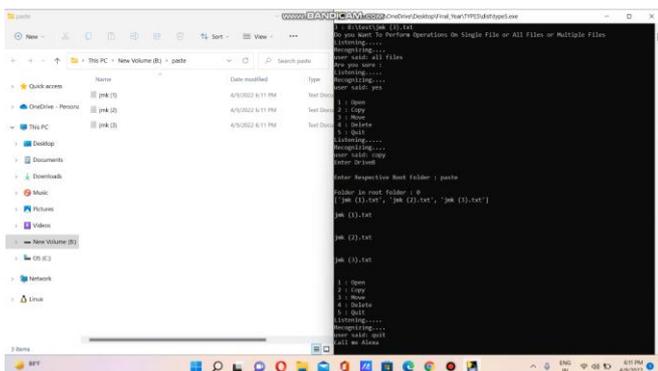


Figure 5. Copy operation completed

After completing operation, “exit” command will take you out from the file implement section to overall operation section.

2. Bookmark File

This feature will allow you to select your favourite file and bookmark it with any voice query or command and it will be permanently saved to your system. Such that whenever you run assistant it will remember the command and as soon as you give voice command to it to open that file it will perform that operation.



Figure 6. Application Bookmarking

After Bookmarking tab, command is ready to use. Just by calling alexa and commanding it to perform operation it can run the specific file as shown below.

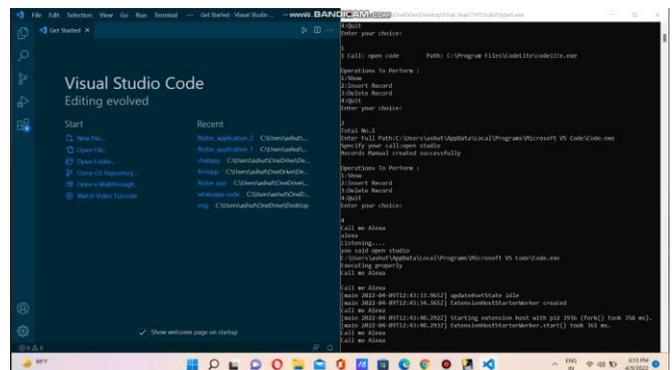


Figure 7. Performing the user’s entered command

5.2 Web Operation Implementation

1. Google Automation

This allows you to perform multiple operations and not only to just search on google. Our system is focused on automating the most of the google processes. Searching and opening google, Opening and closing of tabs, switching tab, Bookmark one or all tabs, opening history, opening download opening bookmark and incognito mode are the things that can be done using our system.

We have shown the example of some operations below:

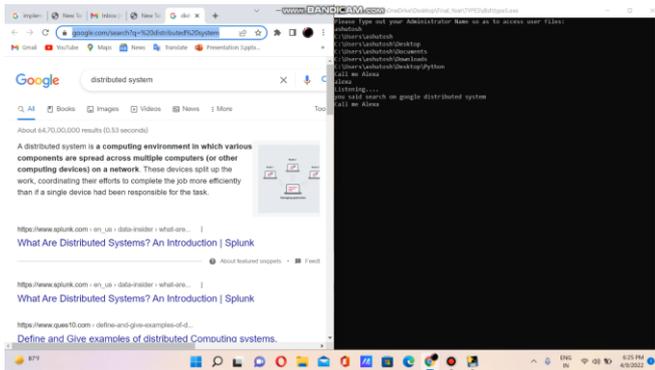


Figure 8. Searching on Google

Figure 9 and 10 deals with the opening new tab with command and closing that new tab again with command

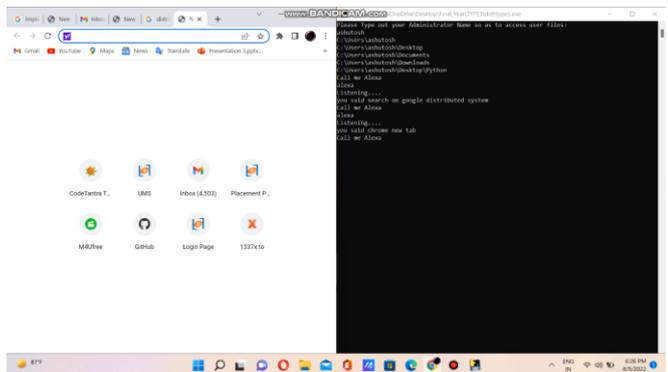


Figure 9. Opening new tab

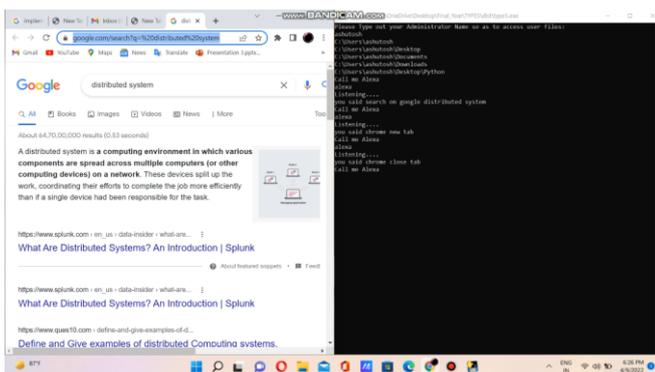


Figure 20. Closing new tab

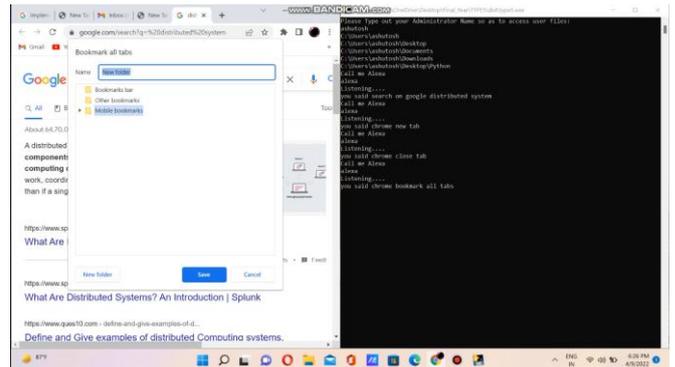


Figure 31. Bookmarking Tabs

Figure 11 and 12 deals with bookmarking all present tabs and opening manager for confirming it.

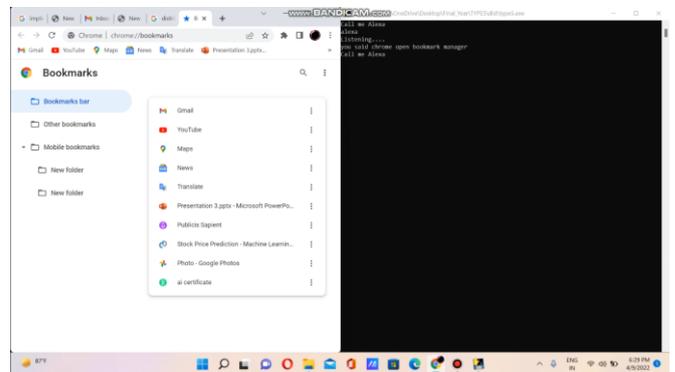


Figure 42. Opening Bookmark Manager

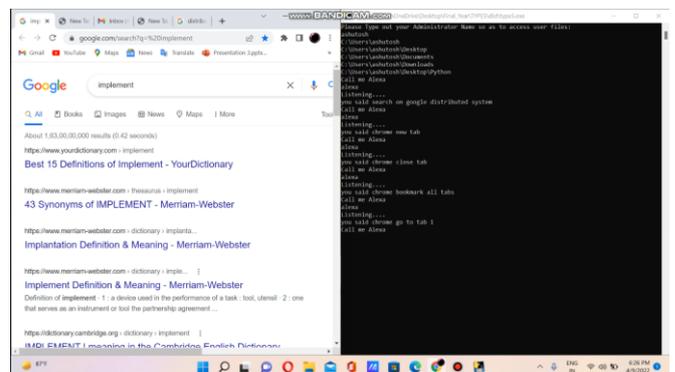


Figure 53. Switching to tab 1

Figure 13 and 14 are related for switching tabs and after that closing all google tabs present



Figure 64. Closing all tabs

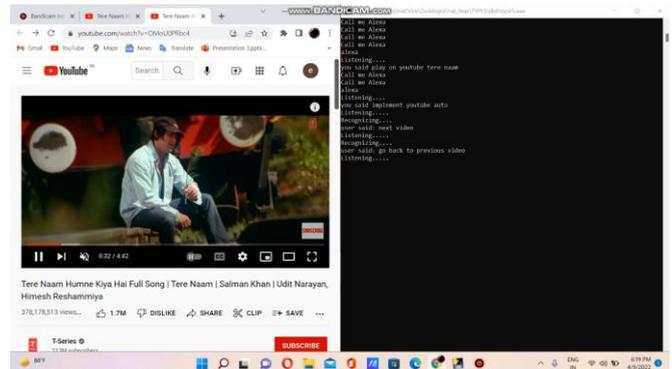


Figure 97. Again going back to previous video

2. Youtube Automation

This will not only just allow to open and play on youtube but will also allow to do the normal executable task like play, pause, mute, unmute, next video, previous video, speed up, speed down and on off subtitle

We have shown the example of some operations below:

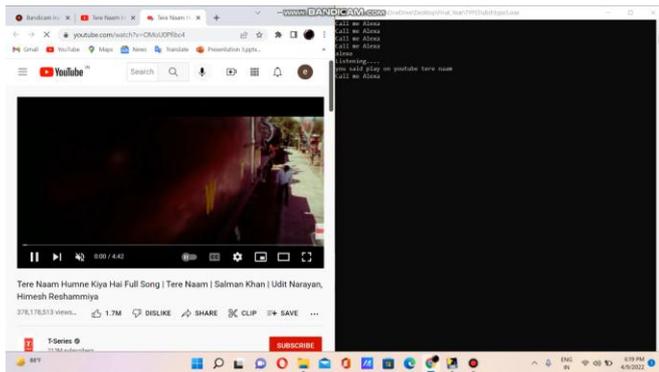


Figure 75. Playing on Youtube

5.3 Other Operations

It allows you to keep up with the other system and web related operations like shutdown, restart the system, testing the net speed, finding important headline, finding particular location/place and telling jokes.

Again some of the operations are shown below



Figure 108. Other Operations

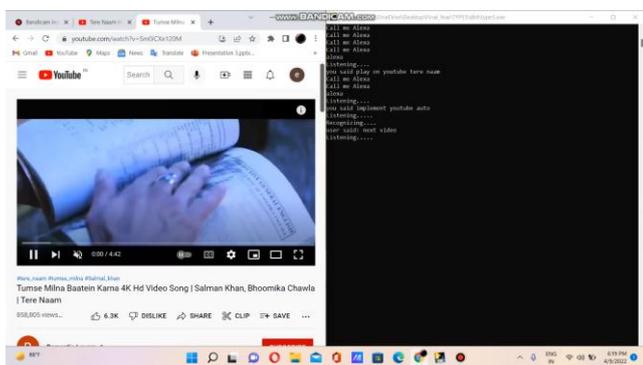


Figure 86. Next Video

6. FUTURE SCOPE

1. For future work, we can investigate how to improve the result of the speech recognition system by doing complex tasks whose commands consist of complex queries (for example: special characters, numbers) by transforming them into easy to remember commands.
2. One idea is to add an eye tracking system which will cover the task of the mouse along with voice recognition for better accuracy of results.
3. Systems can also be incorporated with technologies like machine learning, Neural network, artificial intelligence and IoT. These will help our project to reach a new scale of intelligence.

4. Also the control of Machine can be achieved by synchronizing the automated system or assistant with the mobile so has to control it with the help of mobile.

7. CONCLUSION

This System will help People to carry out most of the work without Handling a computer with a physical keyboard and mouse. As being made available with the help of an executable file, we have ruled out the necessity of the presence of the python language in the system. We have made this assistant or automated system to carry out the file related tasks along with the web search operations. Also this will help most of the physically Challenged people to Make use of Desktop or Laptop by crossing challenging bars, thus making user experience more Friendly and Memorable.

8. ACKNOWLEDGEMENT

Motivation and guidance are the keys towards success. I would like to extend my thanks to all the sources of motivation.

We would like to grab this opportunity to thank Dr. S. D. Sawarkar, principal for encouragement and support he has given for our project. We express our deep gratitude to Dr. A. P. Pande, head of the department who has been the constant driving force behind the completion of this project.

We wish to express our heartfelt appreciation and deep sense of gratitude to my project guide Prof. P. Y. Itankar for his encouragement, invaluable support, timely help, lucid suggestions and excellent guidance which helped us to understand and achieve the project goal. his/her concrete directions and critical views have greatly helped us in successful completion of this work. We extend our sincere appreciation to all professors for their valuable inside and tips during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge them individually. We are also thankful to all those who helped us directly or indirectly in completion of this work

REFERENCES

- [1] Ankush Yadav, A. S. (July 2020). "Desktop Voice Assistant for Visually Impaired" retrieved from International Journal of Recent Technology and Engineering (IJRTE).
- [2] Dr. D. Bhavana, R. V. (DECEMBER 2019). "Advanced Desktop Assistant With Voice" retrieved from International Journal of Scientific & Technology Research.
- [3] Dr. Kshama V. Kulhalli, D. S. (2017). "Personal Assistant with Voice Recognition Intelligence" retrieved from International Journal of Engineering Research and Technology.
- [4] Kumar, L. (2020). "Desktop Voice Assistant Using Natural Language Processing" retrieved from International Journal for Modern Trends in Science and Technology.
- [5] Aditya Tyagi, H. S. (2020). "Desktop Voice Assistant With Speech Recognition" retrieved from International Journal for Research in Engineering Application & Management (IJREAM).
- [6] R.D. Sharp, C. R. (n.d.). "The Watson speech recognition engine." Retrieved from IEEEExplore: <https://ieeexplore.ieee.org/document/604839>