An Analysis on Voice Assistant

Shravankumar Jha

B.TECH Student, Department of Computer Science and Enginnering, Guru Nanak Institute of Technology, Mullana, Ambala, Haryana, India

Abstract: *Voice assistants are programs on digital devices that listen and respond to verbal commands. A user can say, "What's the weather?" and the voice assistant will answer with the weather report for that day and location. They could say, "Tell me a joke," and the assistant will jump into a tale. The user could even say, "play music," and song will be played ! Voice assistants are so easy to use that many people forget to stop and WONDER how they work. How do voice assistants understand us? Is it magic? A complex system of codes? An actual person listening on the other end? The answer is less complicated than you might think. The application works like Siri, Google Assistant etc. The U.I of the application is self explainable and very minimum. It takes voice as input. The system is being designed in such a way that all the services provided by the mobile devices are accessible by the end user on the user's voice command.*

Keywords: siri, alexa, cortana, Google assistant

1. Introduction

The research is focused on natural language interaction with smart voice assistant system on a computer. A voice assistant is a software agent that can perform tasks or services for an individual based on commands or questions. The term "chatbot" is sometimes used to refer to virtual assistants generally or specifically accessed by online chat. In some cases, online chat programs are exclusively for entertainment purposes. Some virtual assistants are able to interpret human speech and respond via synthesized voices. Users can ask their assistants questions, control home automation devices and media playback via voice, and manage other basic tasks such as email, to-do lists, and calendars with verbal (spoken?) commands. A similar concept, however with differences, lays under the dialogue system. Well, I had the similar thought before I started making my very own "JARVIS" Personal Assistant. Though it is not as capable and high as like Amazon's Alexa or Google Assistant, Home or Apple's Siri or JARVIS from Iron Man. Nowadays, People are troubled by typing commands into the computer. Be it procrastination or a busy schedule. Typing is a big obsolete process. The solution to this is that we switch over to an assistant which understands us and do the initial work for us. An assistant is the best replacement for typing commands.

It's named as desktop Voice Assistant "JARVIS" with Voice Recognition Intelligence, which takes the user input in form of user's voice and processes it and return the output in various ways like an action to be performed or the search result is speaked out to the end user. Jarvis is virtual helper. It allows you to get stuff done faster. You can use voice commands instead of performing tasks manually by tapping your finger on the screen a gazillion times "Jarvis" is smart and very well integrated with windows. It can open apps, send messages, make calls, play a specific song on YouTube Music, check the weather, control smart devices, set timers, grab general information, and many other things To use Jarvis on your desktop you first have to enable it. The process is so simple even your anyone can do it who have a basic knowledge about computer, and it won't take more than a minute or two of your time.

2. Methodology

The part where I tell you what the basic requirement for this project are. You'll need **Python 3.6**. We'll be using the pyttsx3package which is a text-to-speech library for Python. The basic reason why we use this is because it works offline. Another basic requirement of this project will be Python's Speech Recognition library. There are other requirements for the project which are listed below; we'll understand them as we go ahead. Inappropriate college description is also conveyed as all terms and conditions of college are not known to students.

Python
Python 3.9.5 is the newest major release of the Python programming language, and it contains many new features and optimizations. There's been 111 commits since 3.9.4 which is a similar amount compared to 3.8 at the same stage of the release cycle. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a

program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

3. Modelling and Analysis

Speech recognition

Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, is a capability which enables a program to process human speech into a written format. While it's commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text one whereas voice recognition just seeks to identify an individual user's voiceSpeech recognition processing Speech recognition is the process of converting human sound signals into words or instructions. Speech recognition is based on speech. It is an important research direction of speech signal processing and a branch of pattern recognition. The research of speech recognition involves many subject areas such as computer technology, artificial intelligence, digital signal processing, pattern recognition, acoustics, linguistics, and cognitive science. It is a multidisciplinary comprehensive research field. Different research areas have emerged based on research tasks under different constraints. According to the requirements of the speaker's way of speaking, these areas can be divided into isolated words, connected words, and continuous speech recognition systems. According to the degree of dependence on the speaker, these areas can be divided into speech recognition systems for the specific person and nonspecific person. According to the size of vocabulary, they can be divided into small vocabulary, medium vocabulary, large vocabulary, and infinite vocabulary speech recognition systems.
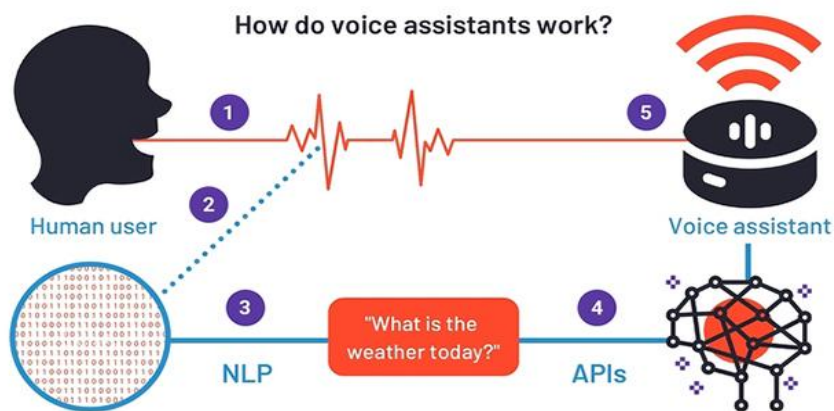
4. Conclusion

Controlled Devices uses Natural Language Processing to process the language spoken by the human and understand the query and process the query and respond to the human with the result. The understanding of the device means Artificial Intelligence needs to be integrated with the device so that the device can work in a smart way and can also control IoT applications and devices and can also respond to query which will search the web for results and process it. It is designed to minimize the human efforts and control the device with just human Voice. The device can also be designed to interact with other intelligent voice-controlled devices like IoT applications and devices, weather reports of a city from the Internet, send an email to a client, add events on the calendar, etc. The accuracy of the devices can be increased using machine learning and categorizing the queries in particular result sets and using them in further queries. The accuracy of the devices is increasing exponentially in the last decade. The devices can also be designed to accept commands in bilingual language and respond back in the same language queried by the user. The device can also be designed to help visually impaired people.

Reference

[1] www.javapoints.com

[2] https://www.python.org

[3] Lee, Nicole. "Google Assistant on the iPhone Is Better than Siri, but Not Much."Engadget (May 17, 2017). https://www.engadget.com/2017/05/17/googleassistant-iphone-hands-on/.

[4] DOUGLAS O'SHAUGHNESSY, SENIOR MEMBER, IEEE, "Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis" proceedings of THE IEEE, VOL. 91, NO. 9, SEPTEMBER 2003

[5] Kei Hashimoto1, Junichi Yamagishi2, William Byrne3 Simon King2, Keiichi Tokuda, "An analysis of machine translation and speech synthesis in speech-to-speech