

PERSONAL VOICE ASSISTANT

Mr. K. Vikram Reddy, Assistant Professor, Department of Computer Science and Engineering, Matrusri Engineering College, Saidabad, Hyderabad, Telangana,

S. Lahari, A. Naveen, G. Sarath Chandra, B.E Scholars, Department of Computer Science and Engineering, Matrusri Engineering College, Saidabad, Hyderabad, Telangana.

ABSTRACT - Digitization brings new possibilities to ease our daily life activities by the means of assistive technology. Amazon Alexa, Apple Siri, Microsoft Cortana, Samsung Bixby, to name only a few were successful in the age of smart personal assistants (spas). A voice assistant is defined a digital assistant that combines artificial intelligence, machine learning Speech Recognition, Natural Language Processing (NLP), Speech Synthesis and various actuation mechanisms to sense and influence the environment. We use different NLP techniques to convert Speech to text (STT), then process the text, convert Text to Speech (TTS), add various functionalities. However, SPA research seems to be highly fragmented among different disciplines, such as computer science, human-computer-interaction and information systems, which leads to 'reinventing the wheel approaches' and thus impede progress and conceptual clarity. In this paper, we present an exhaustive, integrative literature review to build a solid basis for future research. Hence, we contribute by providing a consolidated, integrated view on prior research and lay the foundation for an SPA classification scheme.

Instead of pattern recognition we use nlp techniques to recognize the text which is context based. Operates online as well as offline. Data is Stored in Application itself, reduces Time and Space Complexity.

1. INTRODUCTION

In 21st century, everything is leaning towards automation, may it be your home or car. There is an unbelievable change rather advancement in technology over the last few years. Believe it or

not, in today's world you can interact with your machine. What is interacting with a machine? Obviously giving it some input, but what if the input is not in the conventional way of typing, rather it is your own Voice. What if you are talking to the machine, giving it commands and wanting the machine to interact with you like your assistant? What if the machine is not giving you answers just by showing you the best results but also by advising you with a better alternative. An easy access to machine with voice commands is the revolutionary way of human system interaction. To achieve this, we need to use speech to text API for understanding the input. Many companies like Google, Amazon and Apple are trying to achieve this in generalized form. Isn't it amazing that you can set reminders by just saying remind me to... Or set alarm with wake me up at .. Understanding the importance of this we have decided to make a system that can be placed anywhere in vicinity and you can ask it to help you do anything for you just by speaking with it. In addition to this, you can also connect two such devices through WiFi and make them communicated with each other in future. This device can be very handy for day to day use and it can help you function better by constantly giving you reminders and updates. Why would we need it? Because your own voice is turning into a best input device than a conventional enter key. All the Operating Systems provide plenty of applications and services for users. The Most famous application of iPhone is "SIRI" which helps the end user to communicate end user to mobile with voice and it also responds to the voice commands of the user. Same kind of application is also developed by the Google that is "Google Voice Search" which is used for in

Android Phones. But this Application mostly works with Internet Connections. But our Proposed System has capability to work with and without Internet Connectivity which takes the user input in form of voice or text and process it and returns the output in various forms like action to be performed. Voice controlled home automation systems could offer people a more comfortable lifestyle and simplify ordinary tasks. Voice control within sustainable homes is especially beneficial for people with disabilities, enabling a lifestyle, which were previously impossible. An implementation of voice command systems could have great benefits, also offering help with assistance at the workplace.

A voice assistant is a software agent that can perform tasks or services for an individual, by using the technology of voice recognition. Following voice assistants are a few of which can be found on the market today-Apple's Siri, Amazon's Alexa, Microsoft's Cortana, Google assistant, Samsung Bixby and many more.

A voice assistant is a software agent that performs tasks or services assigned to it by the user through various commands. The software agent which is accessed by online chat is referred as 'chatbot' in software terms which is a part of Virtual agent domain. Voice assistants of the same domain are able to interpret human speech and respond to it. On the other hand, data storage is major issue in day-to-day life. Every smart device has a limited data storage capacity. After the storage is full, managing data is a tedious job. For this given issue, we make use of multiple hard drives which also have a limited capacity. Moving data from different storage devices becomes difficult and is time consuming. Once the surface of the disk is damaged by regular crashes, it leads to the loss of the data stored in the disk. Cloud Storage is a model of computer data storage in which the digital data is stored in various logical pools of disks, but there is a need of an active internet connection for accessing the data stored in the cloud. Several cloud storage services have a specific bandwidth allowance; the additional

charges could be significant and requires more time. Hence creating a file storage system accessed by heterogeneous client devices and integrating this system to voice assistant software agent makes a smart storage device.

Voice control further improves the convenience provided by such devices, and it has already been applied in various systems. For example, the aforementioned driver could control the GPS system in his car without letting go of the steering wheel, while the busy secretary could simply tell her phone to dial a number while she works on an important file. Although, the more technically savvy users may prefer to use such a system because they prefer talking to typing, or simply because it's fun.

2. LITERATURE SURVEY

Voice assistant has a long history with several waves of major innovations. Voice assistant for dictation, search, and voice commands has become a standard feature on smartphones and wearable devices. The study stems from an overlooking literature review in order to present generic knowledge (theory and concepts) about voice control, virtual assistants, fields of use and more. When looking at a number of currently available intelligent programs with natural language processing capabilities, many examples can be found in everyday life filling a variety of roles. The first speech recognition system, named Audrey, was created by Bell Laboratories in 1952. Audrey was rather rudimentary and limited technology wise, understanding only ten digits - spoken by particular people (Pieraccini, 2012). About 10 years later, IBM developed and demonstrated their Shoebox Machine. The device recognized and responded to 16 different spoken words, including all ten digits "0" to "9" as well as calculating commands such as "plus" or "minus" (IBM, 2018). Shoebox Machine recognized and responded to 16 spoken words, including the ten digits from "0" through "9", only in English by a designated speaker. These limitations later proved to be problematic, increasing the scepticism opposing voice recognition. Mid 1970's came the Hidden Markov

Model (HMM) (Rabiner,1989). The HMM considerably altered the development of a feasible speech recognition software. With the help of HMM speech recognition started using a statistical method measuring the probability of unknown sounds being words. Now, the potential to recognize an unlimited number of words became imminent due to the method allowing the number of understandable words go up to a few thousands. These choices of observation distribution in each state of the model allow accurate modelling of virtually unlimited types of data. The first mass accessible voice command system was launched by Apple Inc. as they released the virtual assistant named Siri in 2011 (Bostic, 2013).The intelligent bot Siri can be found as standard on Apple mobile devices now and is considered a core component on these devices. Siri is a personal assistant that uses natural language processing to answer questions and outsource requests to web services that will then be carried out for the user. Similarly to this, the chatbot HAL was created by Zabaware Inc to function as a virtual assistant for users on computers. The bot also uses natural language processing algorithms to converse with the user and take notes from what the user is saying in an effort to organize the data given to it. IBM has invested a large amount of resources into this field and has created Watson, a system developed to compete on the TV show Jeopardy!. This system exhibits the current capabilities of intelligent systems with natural language recognition as it successfully beat the two most successful human contestants of the show. In contrast to these roles is the chatbot Kari who functions as a virtual girlfriend. This system communicates with the user and through using similar methods natural language recognition tries to provoke social conversation with the user. The software aims to give personal companionship and to replicate human interaction as accurately as possible with the assistance of algorithms designed to help the program learn from its inputs.

3. OVERVIEW OF THE SYSTEM

The overall system design consists of following phases:

1. Data collection in the form of speech.
2. Voice analysis and conversion to text
3. Data storage and processing
4. Generating speech from the processed text output

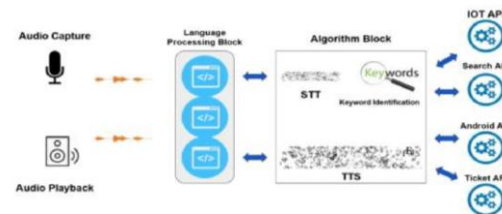


Fig 3.1: System Architecture

In first phase, the data is collected in the form of speech and stored as an input for the next phase for processing. In second phase, the input voice is continuously processed and converted to text using speech to text (STT). In next phase the converted text is analyzed and processed using Python Script and NLP techniques to identify the response to be taken against the command. Finally once the response is identified, output is generated from simple text to speech conversion using text to speech (TTS).

3.1 Problems with Existing System:

Despite the various benefits provided by speech recognition, the system is also plagued with limitations. By implication the development of speech recognition applications also inherits these limitations. The existing Voice Assistants use pattern recognition techniques of python which lack in the context, Lack of accuracy, and misinterpretations, Time, costs and productivity, User accents. They operate only on online mode. They store the data in database servers which leads to increase in Time and Space Complexity. Some of them use cloud to store the data which leads to security issues. Background noise interference is also another daunting problem with speech recognition software.

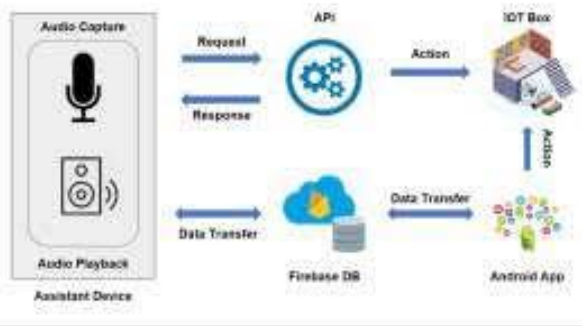


Fig 3.2 Block Diagram of Existing System

3.2 Proposed System

The proposed system of voice assistant will solve some issues of existing system as well introduce new features for better quality and usage. So, let’s have a brief of the new updated version of the voice assistant.

Instead of pattern recognition technique which has been used in previous models, we use Natural Language Processing (NLP) techniques to recognize the text which is context based rather the usual pattern based. This Operates in online as well as offline mode. System application runs on offline mode, whereas web based operations run on online mode. Data is Stored in Application itself, rather than cloud which reduces Time and Space Complexity. It even reduces the economic cost due to reducing high bundles of data usage.

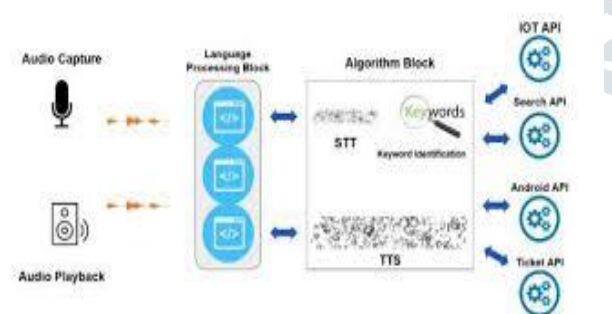


Fig 3.3 Block Diagram of Proposed System

3.3 Process Logic

Personal Voice Assistant is developed as a desktop application with the help of Natural Language Processing which helps to send messages and use various built-in systems based and web based applications using voice commands.

The Voice Assistant performs basic operations such as controlling computer tasks and operations, asking for temperature, humidity, date, time, and year. Adding, reading and deleting notes using voice commands and playing YouTube videos on demand.

The above tasks can be performed using certain methodologies in which each technique has its own functionality and different operations to be performed. Each technique has different process logic to be executed.

Techniques to be implemented:

1. Speech Recognition
2. NLP
3. Threading
4. Scraping

4. IMPLEMENTATION

Offline Speech Recognition:-

The application is able to recognize the voice from the user end without internet connection. The verbal commands are converted to text and the operations take place.

Content Opener:-

The application is able to open files, folders, drives, documents present in the system.

Making Notes:-

The application is able to create, read, write and delete notes.

Search:-

The application is able to search YouTube videos, songs, Wikipedia, google for any information in the internet.

Updates:-

The application gives updates regarding the date, time, day, weather, humidity, temperature.

5. RESULTS



Fig 5.1: Greetings of Voice Assistant



Fig 5.2: Time, Date, day and Year



Fig 5.3: Weather, Temperature and Humidity

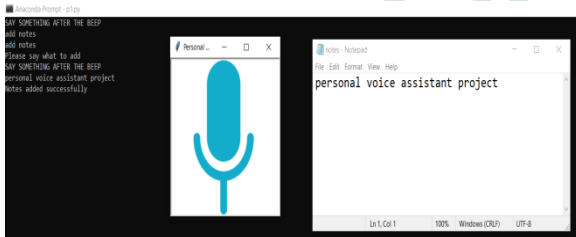


Fig 5.4: Adding Notes

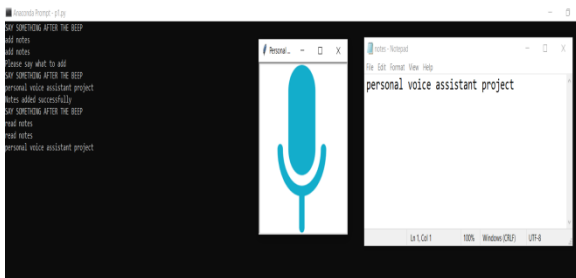


Fig 5.5: Reading Notes

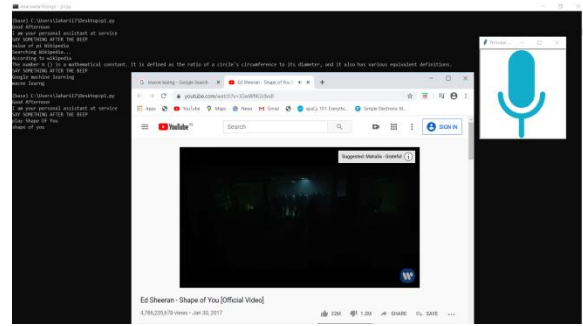


Fig 5.6: Playing Videos in YouTube

6. CONCLUSION

Voice Controlled Personal Assistant System will use the Natural language processing and can be integrated with Machine learning techniques to achieve a smart assistant that can perform action on various applications and will make human life comfortable. The system will have the following phases: Data collection in the form of voice; Voice analysis and conversion to text; Data storage and processing; generating speech from the processed text output. This application will also make life easier for those who are physically disabled and every common user who is fascinated by voice recognition. Academically, raising awareness for systems like this for students can give them better understanding of topics like Artificial Intelligence, Neural Networks, Natural Language Processing, Machine Learning and Human Computer Interaction and also how to improve user experience in application development. The formulated solution is able to process voice commands offline allowing users to cut down on the cost of data bundles. This also helps to make it faster in comparison to alternative applications like Apple's Siri, Google assistant, etc. Moreover, the solution is capable of carrying out a variety of tasks with ease such as telling the date and time, playing music/videos, making phone calls ,finding weather, temperature, googling information etc.. This paper can also act as a prototype for many advanced applications.

7. FUTURE ENHANCEMENTS

Based on the survey we recommend that the application should be developed which accomplishes the desire of different users. The main reason that the user wants to use the voice assistant is to make their life easier, so by implementing the below mentioned features the user can be facilitated.

1. Developing for different languages and different accents.
2. Portability for any environment.
3. Voice authentication technology can be implemented for more security.
4. Chatbot implementation requires corpus.
5. Dialogue flow needs stack with neurals
6. Deploy on web using flask or Django
7. Deploy on cloud uses amazon ec2, Heroku.
8. NLP features such as finding entities, topic modelling.

8. REFERENCES

1. 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
2. Nil Goksel-Canbek2 Mehmet EminMutlu, "On the track of Artificial Intelligence: Learning with Intelligent Personal Assistant" International Journal of Human Sciences, 13(1), 592-601. doi:10.14687/ijhs.v13i1.3549.
3. Easwara Moorthy, A., Vu, K.-P.L.: Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. International Journal of Human-Computer Interaction 31, 307–335 (2015)
4. Tsiao, J.C.-S., Tong, P.P., Chao, D.Y.: Natural-Language Voice-Activated Personal Assistant, United States Patent (10), Patent No.: US 7,216,080 B2 (45), 8 May 2007
5. Lopez, G., Quesada, L., Guerrero, L.A.: Alexa vs Siri vs Cortana vs Google assistant: a comparison of speech-based natural user interfaces. Conference Paper, January 2018
6. Kepuska, V., Bohouta, G.: Next generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In: IEEE Conference (2018)
7. Gong, L.: San Francisco, CA (US) United States US 2003.01671.67A1 (12) Patent Application Publication c (10) Pub. No.: US 2003/0167167 A1 Gong (43) Pub. Date: 4 September 2003 for Intelligent Virtual Assistant
8. Sumitkumar Sarda, Yash Shah, Monika Das, Nikita Saibewar, ShivprasadPatil,"VPA: Virtual Personal Assistant" Published in 2017 International Journal of Computer Applications (0975 – 8887) Volume 165 – No 1.
9. Grabianowski, E. (2011). How Speech Recognition Works. Retrieved February 2016, from How Stuff Works: <http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition1.html>
10. Nagesh Singh Chauhan,"Build Your First Voice Assistant",March,2019 <https://towardsdatascience.com/build-your-first-voiceassistant>.