International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

VIRTUAL VOICE ASSISTANT

Ravikumar N R[1], Prateek C[2], Sathvik Bhandar[3], Rahul Kumar[4], Mayura D Tapkire(Assistant Professer)[5]

[1-5]*Department of Information Science and Engineering National Institute of Engineering, Mysuru, India.*

Abstract - *In the 21st century, everything is automated including the things we use daily washing machines, dishwashers, refrigerators, bus doors, air condition systems, turning everything in a single click, etc. In this fast-moving the present study proposes the newer concept of voice-controlled devices that recognizes one's voice; process the request and assigns the time and date of the appointment based on the request with details as name of the person; date; time and other related information. We need to develop devices with in-built voice recognition which has the ability to recognize the voice even in crowded surroundings with just only one form of interaction between the device and the human is the Voice. The device will capture the audio through the microphone of the device and process the query made by the human and reply to the human with the appropriate results. For Example if you ask the device to change the wallpaper of your Personal Computer it will change the wallpaper by downloading wallpaper from a website and changing the wallpaper. It can also guide you the traffic between source and destination and also auto-suggest lesser traffic and time routes.*

Key words: automated, dishwashers, recognizes, crowed, website, destination.

INTRODUCTION

This Voice recognition technology is evolving rapidly and is expected to become not only the default input form for smartphones, but also for cars and other home appliances such as TV and fridge. Due to the unique features associated with voice input, including an implicit verbalization of commands, privacy and acceptability issues may affect the usage and adoption of voice-bases.

Several researchers have been interested in the recognition of human activities in recent years. In this project, we propose a voice recognition system that recognizes human activities through a deep learning algorithm. Voice is essentially a mode of communication that lets users communicate with each other. Voice Recognition, also known as Automatic Speech Recognition (ASR), identifies spoken words and phrases and translates them to a readable computer format. It takes the input of the user in the form of a voice or text and processes it and returns the feedback in different ways, such as the action to be done or the search result to the end user. Hence, there is also the additional challenge of making out spoken words from noise in the audio. Modern mobile technology has been very useful to consumers as it provides consumers with access to various apps and resources from anywhere in the world. Some of the most used mobile Operating Systems are Android, Apple, Windows, Blackberry, etc. These Operating System provides different services to the user. Most systems allow the user to train the software to understand their voice so that it can translate speech to text more precisely. Google has created "Google Assistant" and iPhone has created "Siri" to help users respond to their voice commands in an effective manner. Mobile App Voice Assistant helps users to communicate with AI in a very meaningful way. Mobile App Voice Assistant helps users to communicate with AI in a very meaningful way. In using natural language processing and machine learning, Amazon's Echo app has enhanced the engagement of individuals with AI technology.

Speech recognition is an alternative to keyboard typing. Simply put, you're talking to the machine and your words show on the screen. It has been developed to provide a simple way to write on a computer and can support people with a range of disabilities. It is helpful for users with hand disabilities who often find it though, impossible or painful to type. Voice-recognition apps can also support people with spelling problems, including those with dyslexia, since well-recognized words are almost always pronounced correctly. Scientists have used text generated online by people to train voice assistants to listen and respond to our requests in a more natural and meaningful way. Voice assistants will decipher questions that have been phrased in a variety of different ways and interpret what the user is most likely to want.

2. LITERATURE REVIEW

This research could be is a chunk of a bigger project concerning virtual voice assistant briefed by theories in human machine interaction. Moreover speech recognition has a brief history with numerous waves of innovations. Voice recognition for dictation, hunt and voice command has become vital feature on personal devices: like wearable devices and smartphone's. This system was developed as a humanoid application that confirms the necessity of language rework that sends messages and also use build-in application by processing the commands given by user to the system. Importantly smartphone gadget was way quicker followed by other wearable devices; so, many arrived to introduce in-voice virtual voice assistant with the importance of adopting and applying multiple smart technologies. This system has

some basic features and most importantly mailing and secondly calendar, where user has the privilege to mail and able to create their required event by providing voice commands. For instance, if we use artificial intelligence we can are able to turn off the lights without the instruction given by the user. Almost, Everyone has some knowledge about trending voice assistant like cortana for windows, and Siri for apple users, this virtual voice assistants aren't as brainy and intelligent as Ironman's Jarvis which appear in the superhero movie, but the intended actions are almost similar by virtual voice assistant. It's like you need a ask question, and within a few fraction of seconds you will get an answer. It's just give a command and get result.

Here are some amazing Features of virtual voice assistant:

Open any website in the browser: If any user needs to open any website they just need to voice out "open nameofwebsite.com" or "open website.org". consider example:"open xhsj.com" or"hey requesting you please open zzz.com" .

Plays song on VLC media player: Ask voice assistant to play a desired song in VLC media player :For instance user will ask voice assistant"can you please play me a song", whereas bot will ask"what song shall a pay Sir/Madam?" and voice assistant will transfer the required music to youtube, which is present in your local drive and it will stream the searched content in VLC media player , however if user plays any new song, previously downloaded music will be automatically deleted .

Scan the Headlines: raise voice assistant to scan out daily headlines from news connected application, where user has the privilege to select the interested topics of his/her own alternative.

Send Email: If user had prompted the word email in his/her commands then the voice assistant will ask user for recipient, If user response is abc, then assistant will use phones library for search user data and then it is directed to email with recipient name on it.

Tells you the current time: Using voice assistant users can ask the current time. For instance: "whats time right now" then assistant will report you the current time as per your timezone.therefore"current time is 1.14p.m."

Keep reported about the weather and temperature of any world: Voice assistant can report weather for the day and it also can give as the minimum and maximum temperature of any city across the world. User must just give commands like "what is the current weather in Mysore "or tell me the current weather in "India": you will be getting results within fraction of seconds.

Answer your Desirable questions: Ask voice assistant some interesting facts or the new facts, solve some basic mathematical problems or we are also able to ask a joke.

3. METHODOLOGY

3.1 General Structure

Considering overall research, voice application will be used in following three ways: Firstly, command to the computer whereas secondly, to input information the computer, finally for communication with other people. In this section we will be discussing general components for voice application. As seen in Figure[1],voice will be divide into four different parts: front-end interface, end users, voice recognition System finally dictionary and text file database. Each section is explained as follows:

Front-End Interfaces

In front-end-interface, user will be having direct access to the interface and communication users by providing Input and Output with graphics designs and icon-based menu. It receives user prompt input voice and in return delivers users with a voice recognition system to detect voice inputs, and usually generates feedback of voice to users, after completion of commands by several other functions of the system.

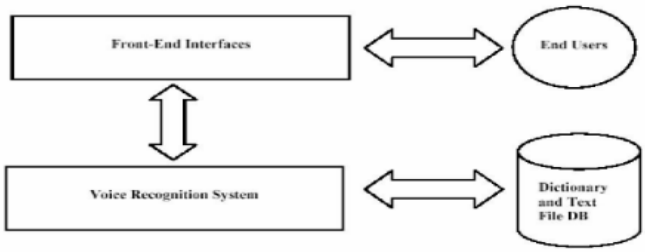


FIGURE 1: General Structure of Voice Application.

End Users

Basically end users refers to device users. They will be using this devices for communication and feedback of voice with the use of application, and moreover end users are those who will be using this application with there personal devices like mobile and laptop users.

3. Voice Recognition Systems

It is the heart of a Voice application system, which has ability to understand voice input given by user, and make application work in a efficient way and generating voice feedback to the user. This system is an important component for user as a gateway to use his or her voice as a input component. In a Nutshell, for clearly understanding user voice command and to get feedback from the system, we should consider voice recognition system contains all the process by which application system directs for building speech signals to text data and few form of important meaning of speech.

3.1 Dictionary and Test files Database

Providing the type of the device and requirements for the user, the system application needs to support few exact input types or by providing peculiar voice feedback. Whereas for language type, the system application can give additional explanations to the users or it may provide functions based on the files in the database in extension, system requires to install an additional text file database to add and update application in few different cases.

3.2 System Architecture

The total design consists of these phases:

1) Collection of data which is in speech format.

2) Analyse the voice and convert it to text.

3) storing the data and processing it.

4) Speech generation from the text output that is processed.

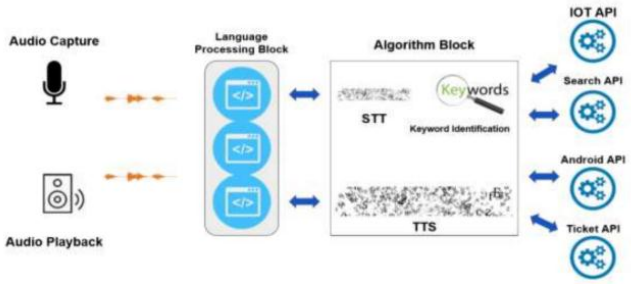


FIGURE 2: System Architecture of Voice Controlled Personal Assistant

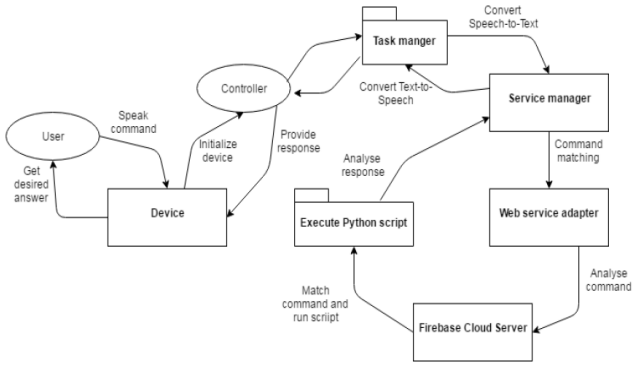
The data that is collected in the speech form is stored and used as input for next phase of the process. In next phase, the input which is given in the form of voice is processed continuously and is converted into text by using STT. In third phase, the text which is converted, is analysed by Python Script which processes it and identifies the action to be taken for the command. In the last phase, after the action to be taken is identified, output will be obtained from text to speech conversion using TTS.

3.3 Data Flow Sequence

3.3.1 Initialize Device:- Device initialization does whatever steps are necessary to get a system into a working state. It set the unit in motion by calling its name. The process is specific for every device, there are no magic values that would initialize any device that you come across.

3.3.2 Service Manager:-It helps in Command analysis and match with Web service adapter and cloud server. The semantic description is used to match the command to the Web service adapter.

FIGURE 3. Data Flow Sequence



3.3.3 Task Manager:- The Speech-to-Text and Text-to-Speech transfers are done by the task manager. It provides semantic description of what was spoken, which is passed into a service manager.

3.3.4 Device:- For speech output, a natural language (NLG) component and a text-to-speech (TTS) component is used.

3.3.5 Firebase Cloud Server:- Firebase is a Backend-as-a-Service - BaaS - which hosts a plethora of APIs to perform certain tasks. We analyse the command given by the web service adapter and further it matches the command and run the script with python script.

3.3.6 Device:- For speech output, a natural language (NLG) component and a text-to-speech (TTS) component is used.

3.3.7 Web Service Adapter - It calls the correct Firebase Cloud Service adapter based on the command.

3.3.8 Execute command : Run the Respective python script after you have found for the match for the given order.

4. APPLICATIONS

There are a wide variety of services which are provided by the voice-enabled devices which range from simple commands like providing information about the weather of a place, general information from Wikipedia, movie rating from IMDB, setting an alarm or reminder, creating a to-do list and adding items to the shopping list so that we don't forget when we go shopping. It can also read books for the user or else play music from any streaming services depending on the device provider or user preference. It can also play videos from YouTube or else from any streaming services. In a recent study, voice assistants are also being used to assist public interactions with the Government and also a decrease of 30% work-

load on humans when voice assistants are used in call-centers.

5. LIMINATIONS

The devices which use the human voice for interacting with the device use single commands as input for the device they usually consist of single phrases. When commands become ambiguous, the resulting actions can be misunderstood by the devices. There is only one-way communication between the user and the device because the device cannot talk back for clarification. The applications on the devices cannot reply back with the state of the process whether it is ongoing or completed. There are many cases where only specific tasks are allowed to be done by the voice-enabled devices because of stove top cannot/should not be turned on when there is no one in the kitchen/house. The devices can not integrate context data. They can not log any history about the queries made but they can be trained to learn about the user behavior and learn about the user's usage statistics and give a recommendation to the user according to the time, place, or by any other calculated parameters.

6. CONCLUSION

Voice-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.

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. Kei Hashimoto1, Junichi Yamagishi2, William Byrne3 Simon King2, Keiichi Tokuda, "An analysis of machine translation and speech synthesis in speech-to-speech translation system" proceedings of 5108978-1-4577-0539-7/11/$26.00 ©2011 IEEE.

3. Hirschberg, Julia, and Christopher D. Manning. "Advances in Natural Language Processing." Science 349 (6245): 261–266. doi:10.1126/science.aaa8685.

4. Moore, Clayton. "The Most Useful Skills for Google Home." Digital Trends (May 3, 2017).https://www.digitaltrends.com/home/google-home-most-useful-skills/.

5. 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/google-assistant-iphone-hands-on/.

6. "Accurate and compact large vocabulary speech recognition on mobile devices," in INTERSPEECH.2013, pp. 662–665, ISCA.

7. "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in ICML, 2006, pp. 369–376.

8. "Lattice-based optimization of sequence classification criteria for neural-network acoustic modeling".

9. R. Rosenfeld, D. Olsen, and A. Rudnicky, "Universal speech interfaces," in Proceedings of the International Conference on Spoken Language Processing, Beijing, China: 2000.

10. T. Hornstein, "Telephone voice interfaces on the cheap," in Proceedings of the UBLAB '94Conference, 1994.

11. F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS Q., vol. 13, no. 3, pp. 319–340, 1989.

12. M. Fishbein and I. Ajzen, Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research, Addison- Wesley Publishing Company, Inc.: Reading, 1975.

13. R. Want and B. N. Schilit, "Interactive Digital Signage," in Computer,
vol. 45, no. 5, pp. 21-24, May 2012.

14. Dempsey P. The teardown: Google Home personal assistant//Engineering & Technology. – 2017. – T. 12. – No. 3. – C. 80-81.[2] Chung H. et al. Alexa, Can I Trust You? //Computer. – 2017.– T. 50. – No. 9. – C. 100-104.

15. Arriany A. A., Musbah M. S. Applying voice recognition technology for smart home networks //Engineering & MIS(ICEMIS), International Conference on. – IEEE, 2016. – C. 1-6.