AI Based Virtual Assistance on Raspberry PI

Kamran Quamar[1], Palash Pramod Dhongde[2], Pawan Sanjay Thakare[3],
Ritik Vijaykumar Yadav[4], Sayyed Tarique Hasan A Rahim[5]

Students, Department of Computer Science & Engineering[1,2,3,4,5]
Anjuman College of Engineering and Technology, Nagpur, Maharashtra, India
kamranquamar2579@gmail.com[1], palashdhongde08@gmail.com[2]
pawanthakare899@gmail.com[3], yritik866@gmail.com[4], sayyedtarique4@gmail.com[5]

Abstract: *Artificial intelligence (AI) technologies are one of the new technologies with new complicated features that are emerging at a fast pace. Although these technologies seem to be extensively adopted, people do not intend to use them in some cases. Technology adoption has been studied for many years, and there are many general models in the literature describing it. However, having more customized models for emerging technologies upon their features seems necessary. In this study, we developed a conceptual model involving a new system quality construct, i.e., interaction quality, which we believe can better describe adoption of AI-based technologies. In order to check our model, we used a voice assistant system (VAS) technology as an example of this technology, and tested a theory-based model using a data set achieved from a field survey. Our results confirm that interaction quality significantly affects individual's trust and leads to adoption of this technology. Voice assistants are software agents that can interpret human speech and respond via synthesized voices. Apple's Siri, Amazon's Alexa, Microsoft's Cortana, and Google's Assistant are the most popular voice assistants and are embedded in smartphones or dedicated home speakers. 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 commands. This column will explore the basic workings and common features of today's voice assistants. It will also discuss some of the privacy and security issues inherent to voice assistants and some potential future uses for these devices. As voice assistants become more widely used, librarians will want to be familiar with their operation and perhaps consider them as a means to deliver library services and materials.*

Keywords: Human Computer Interaction; Internet; Libraries; Software Agents; Speech Recognition; Voice Assistants

I. INTRODUCTION

An intelligent virtual assistant (IVA) or intelligent personal assistant (IPA) 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 commands. A similar concept, however with differences, lies under the dialogue systems.[components, incorporating the applicable criteria that follow.

There have been some very good innovations in the field of speech recognition. Some of the latest innovations have been due to the improvements and high usage of big data and deep learning in this field. These innovations have been attributed to the technology industry using deep learning methods in making and using some of the speech recognition systems, Google was able to reduce word error rate by 6% to 10% relative, for the system that had the word error rate of17% to 52%..

First part is responsible for converting numbers and abbreviations to a written word format. This is also referred to as normalization of text. Second part involves the signal to be processed into an understandable one.

Text to speech conversion is the process of converting a machine recognized text into any language which could be identified by a speaker when the text is read out loud. It is a two step process which is divided into front end and back end.

Speech Recognition is the ability of a machine, for instance a computer to understand words and sentences spoken in any language. These words or sentences are then converted to a format that could be understood by the machine. Speech recognition is basically implemented using vocabulary systems. A speech recognition system may be a Small Vocabulary-

many user system or a Large Vocabulary- small user system

II. EXISTING SYSTEM

To process the virtual assistance (VI) the existing system uses mobile phones to trigger , which is limited and resource dependent and quite economically ineffective. Tech Giant company like Google,Amazon.Microsoft Azure are some of the company providing there best services as a voice assistance but there are some privacy issue concern with it as they store voice data to there local cloud storage, which is potentially for data to be stolen ,leaked or used to incriminate people[].

The market value of voice assistance is quite high in the recent year apart from daily home usage now it mostly use in industrial task, business , education and health sectors[], which is quite helpful to handle some of the complex task that human took much of their hours to handle

To avoid all these resource dependency and privacy issue in the existing system we build our project using simple python and deploy it to particular hardware called raspberry pi which make our project mobilize some of the product widely available in our existing system

2.1 Amazon Alexa

Also known as alexa, It is cloud based service which uses natural language understanding techniques to operate the device like amazon echo Echo, Echo Plus and Echo Dot.[]

It also uses automatic speech recognition, natural language processing techniques and some other forms which make artificial intelligence(AI) systems to perform actions.

Amazon generally use seven microphone and noise cancellation mechanism to which enable there device to wake using wake word ("Alexa") after successfully capture the voice data it will automatically sent to amazon to amazon computer to process

2.2 Bixby

It is an AI virtual assistant by Samsung electronics which is powered with s.voice[] technology to make the voice assistance work, it is a context awareness that trigger the right operation based on request.

It can also understand natural language that even if information is incomplete it can perform the action accurately[]. unlike other assistance it have also wake word ("Hi Bixby) that allow a lot of internal system to work that handle the complicated task like taking selfies, or installing an app from Playstore with a voice command

2.3 Cortana

It is a digital assistance by microsoft which uses microsoft azure service to process AI voice assistance services fabric platform that resemble to hyperscale microservices based application, featuring easily managed seppertive systems to enrich customer satisfaction , so it is reliable software[].

Cortana team developed a software to make the problem simple so it can handle critical operations such as data feed including traffic , weather information , geographical data. It uses the Bing search engine to perform tasks such as setting reminders and answering questions for the user. Cortana is currently available in English, Portuguese, French, German, Italian, Spanish, Chinese, and Japanese language editions, depending on the software platform and region in which it is used[].

2.4 Google Assistance

Google Assistance is AI powered which is widely available in mobile phone and smart home devices, which enable the user to search any thing on their demand which actually enabled using keyword or wake word (Hi google). It uses the natural language processing that convert speech into sound. It first record the speech as interpreting sound takes a lot of computational power then it sent google server to get out analysis more efficiently

It break the word into individual sound and interact with database which combine word pronunciation and find the word which most closely corresponding to the combination of individual sound after the identifies the key word it create sense of task occur densely like "how can i help you".

Then it send back those request to our device and perform two way conversation.

2.5 Siri

It is a voice assistance develop by apple, it work on watchOS, macOS, tvOS, and audioOS operating systems. It uses voice queries, gesture based control, focus-tracking and a natural-language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Internet services.

When user give any request siri record the frequency and sound waves from our voice and translate it into code it break down the code to understand particular pattern,phase and keyword.the data is determine by by algorithm that traverse through combination of sentence to determine inputted phrase means[]

The algorithm is capable enough to determine the exact meaning of any word, sentence and expression context of any sentence. Once siri determine the request, it start to handle that particular task that need to carried out , at also determine weather or not information need to be accessed within phone's data banks or from online server. Siri is then able to craft complete and cohesive sentences relevant to the type of question or command requested.

With continued use, it adapts to users' individual language usages, searches and preferences, returning individualized results. There's a huge amount of work in Siri that can predict what you're getting at based on keywords that you use, as well as your general habits and language choice. it is designed to adapt to your individual preferences over time and personalize results.

2.6 chatbot

It is a computer programme that simulates a natural human conversation. user can be communicate via chat interface or by voice like they interacting with real human.[]

It is very useful for customer service, request routing or information gathering, medical sector and other e-commerce platform, education, entertainment, finance, news and productivity by helping people by automate conversation and interact with them through messaging platforms.[]. It require continuous testing and turing

A chatbot or chatterbot is a software application used to conduct an on-line chat conversation via text or text-to-speech, and used in dialog systems for various purposes including customer service, request routing, or information gathering. While some chatbot applications use extensive word-classification processes, natural language processors, and sophisticated AI, others simply scan for general keywords and generate responses using common phrases obtained from an associated library or database. There are two main types of chatbots.

- **Data-driven and predictive chatbot:** It is also referred to as virtual assistance or digital assistance, and also task oriented chatbots which is sophisticated, interactive and personalized. It apply predicative intelligence and analytics to enable personalization based on user profile. To create such a powerful system it should contribute good customer experience [chatbotsjournal] by avoiding irritation and make chatbot connects seamlessly
- **Task-oriented chatbots:** It is a single-purpose programs that focus on performing one function. which use NPL and also MLusing rules, NLP, and very little ML,. Interactions with these chatbots are highly specific and structured and are most applicable to support and service functions—think robust, interactive FAQs. Task-oriented chatbots can handle common questions, such as queries about hours of business or simple transactions that don't involve a variety of variables. Though they do use NLP so end users can experience them in a conversational way, their capabilities are fairly basic. These are currently the most commonly used chatbots.

III. PROPOSED SYSTEM

3.1 Working

The proposed system is a voice assistant that uses speech recognition to recognize human speech and deliver relevant results in speech form. This system provides hands-free user experience and is deployed on a mini computer called Raspberry Pi. This device acts as a personal assistant that can help the user in everyday tasks. This device is portable and suitable for usage in workplaces, home etc.
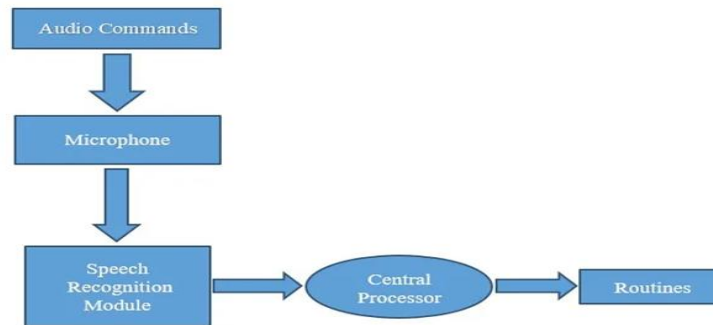
3.2 Module 2: Text to speech

This module receives the output from the Query Processing System and converts the string to the voice for full user-to-user interaction. Text-to-Speech, in particular compared to text-based confirmation, is important to make the virtual assistant humane. This module will contain the

- **Python Backend:** At backend the python gets the output from speech recognition and after that it identifies whether the command is a system command or a browser command. The output is sent back to the python backend to give desired output to the user.
- **API calls:** An API, or Application Programming Interface, is a server that you can use to retrieve and send data using code. APIs are most commonly used to retrieve data
- **Content Extraction:** It is the act or process of retrieving data out of (usually unstructured or poorly structured) data sources for further data processing or data storage (data migration). The import into the intermediate extracting system is thus usually followed by data transformation and possibly the addition of metadata prior to export to another stage in the data workflow.

3.3 Module 3: Interpret Commands

The interpreter operates somewhat like the Unix shell: when called with standard input connected to a tty device, it reads and executes commands interactively; when called with a file name argument or with a file as standard input, it reads and executes a script from that file.

IV. FUTURE SCOPE

The only major drawback in any speech recognition system is its dependency on an active internet connection. It is easier to recognize speech through the internet yet at times or at remote places where a reliable connection may not be available all the time, the device becomes difficult to be used. To overcome this, offline speech recognition systems should be implemented. It will be a great breakthrough when it comes to independent stand-alone home automation devices or virtual assistants because most of the Assistants are connected to the internet

Another important improvement which can be done is the addition of native languages. It is a hectic process yet it will be useful to all the people who do not know English. It makes this device usable by almost all of the people in the world. By the addition of native languages, the device becomes much more user friendly and easily accessible.

Machine Learning should be implemented completely into this system. By that way, the system will be able to learn new processes by itself and adapt to the user based on its past experiences. This makes it easier for the user to interact with the assistant as well. Memory can be improved and let the system store the new information gathered from the user and use it in the future if required. Much more modules can be created based on the necessities.

V. CONCLUSION

The existing voice recognition systems basically run by only a fixed set of commands which makes it a stereotypic functioning. The user loses interest as time moves on. To break the stereotype and to overcome all the issues and problems in the existing system, this proposed system takes in a dedicated python script which completely is interactive not with a fixed set of commands but with dynamically changing responses.

This kind of approach makes the user feel personal while using the system. The software and hardware implemented in this system are mostly open-source and inexpensive when produced in huge quantities will reduce the cost of production very much which ultimately becomes affordable to almost everyone.

REFERENCES

[1]. Jiang, J.; Hassan Awadallah, A.; Jones, R.; Ozertem, U.; Zitouni, I.; Gurunath Kulkarni, R.; Khan, O.Z. Automatic online evaluation of intelligent assistants. In Proceedings of the 24th International Conference on World Wide Web, 18–22 May 2015; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2015; pp. 506–516

[2]. Piyush Vashishta, Juginder Pal Singh, Pranav Jain and Jitendra Kumar, "Raspberry PI based voice-operated personal assistant", International Conference on Electronics And Communication and Aerospace Technology ICECA, 2019.

[3]. Salvatore Gaglio, Giuseppe Lo Re, Marco Morana, and Claudio Ruocco (2019), Smart Assistance for Students and People Living in a Campus, IEEE Computer Society.

[4]. Subhash S, Prajwal N Srivatsa, Siddesh S, Ullas A, Santhosh B (2020), Artificial Intelligence-based Voice Assistant, IEEE.

[5]. Rahul Kumar, Garima Sarupria, Varshil Panwala, Smit Shah, Nehal Shah (2020), POWER EFFICIENT SMART HOME WITH VOICE ASSISTANT, IEEE – 49239.

[6]. Jianliang Meng, Junwei Zhang,Haoquan Zhao (2012), Overview of the Speech Recognition Technology, IEEE.

[7]. Canbek, N.G.; Mutlu, M.E. On the track of artificial intelligence: Learning with intelligent personal assistants. J. Hum. Sci. 2016, 13, 592–601.

[8]. Bill Lubanovic, Introducing Python :Modern Computing in Simple Packages, 2nd Edition, O'Reilly Media, 2014.

[9]. Chelba, Ciprian, et al. "Large scale language modeling in automatic speech recognition." arXiv preprint arXiv:1210.8440 (2012).

[10]. S. Tulshan and S. N. Dhage, "Survey on Virtual Assistant: Google Assistant Siri Cortana Alexa:4th 2018" in Revis. Sel. Pap. Surv. Virtual Assist. Google Assist., January 2019.