AI Voice Assistant (Casper)

A PROJECT REPORT

NAME OF THE CANDIDATE(S)

Kamya Brata Debnath (20BCS3463)
Omkar Singh (20BCS3491)
Harsh Raj (20BCS3381)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE ENGINEERING

Chandigarh University

MAY 2022

BONAFIDE CERTIFICATE

Certified that this project report **"….AI Voice Assistant (Casper)…."** is the bonafide work of " ..... **KAMYA BRATA DEBNATH , OMKAR SINGH , HARSH RAJ ……"** who carried out the project work under my/our supervision.

Gursimran Bakshi
SIGNATURE

GURSIMRAN BAKSHI
SUPERVISOR

Project Teacher

CSE

Puneet Kumar
SIGNATURE

PUNEET KUMAR
HEAD OF THE DEPARTMENT

CSE

Submitted for the project viva-voce examination held on 19 May 2022

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

I would like to express my gratitude to my teacher (Gursimran Bakshi) for providing support and guidance. I got to learn a lot more about this project which will be very helpful for me.

TABLE OF CONTENTS

INTRODUCTION

In recent times only in the Virtual Assistants we can experience the major changes, the way user interacts and the experience of user. We are already using them for many tasks like switching on/off lights, playing music through

streaming apps like Wynk Music, Spotify etc., This is the new method of interacting with the technical devices makes lexical communication as a new ally to this technology.

The concept of virtual assistants in earlier days is to describe the professionals who provide ancillary services on the web. The job of a voice is defined in three stages: Text to speech; Text to Intention; Intention to action; Voice assistant will be fully developed to improve the current range. Voice assistants are not befuddled with the virtual assistants, which are people, who work casually and can therefore handle all kinds of tasks.

Voice Assistants anticipate our every need and it takes action, Thanks to AI based Voice Assistants.

AI-based Voice assistants can be useful in many fields such as IT Helpdesk, Home automation, HR related tasks, voice based search etc., and the voice-based search is going to be the future for next generation people where users are all most dependent on voice assistants for every needs.

In this proposal we have built the AI-based voice assistant which can do all of these tasks without inconvenience.

Literature survey

Voice control is a major growing feature that change the way people can live. The voice assistant is commonly being used in smartphones and laptops. AI-based Voice assistants are the operating systems that can recognize human voice and respond via integrated voices. This voice assistant will gather the audio from the microphone and then convert that into text, later it is sent through GTTS (Google text to speech). GTTS engine will convert text into audio file in

English language, then that audio is played using play sound package of python programming Language.

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 andlay 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.

Design flow/Process

1. **Time:** For the purpose to complete this project we have been given three months aggregate, till the due date to reach we would have our final project with us. The project completion is well defined and it will be completed before the due date. All the necessary features and requirements will be completed in order for proper working of the project. The project is currently in the testing stage and we are making some necessary changes for making the project more user friendly to the user. The project is tested by project members which will then be discussed and will rely on to work on some solutions in order to deliver the project before the due date. The project documentation is also getting reviewed up to date. Once the project outcomes are

documented and the necessary services and objectives are set out to accomplish then we are finally getting closer to the successful project completion.

2. **Cost:** Cost is considered as one of the most important constraints which are required for every project, luckily our project is software dependent which is available free of cost for the developers which is python and its libraries, so, our project is not bounded by cost constraint.

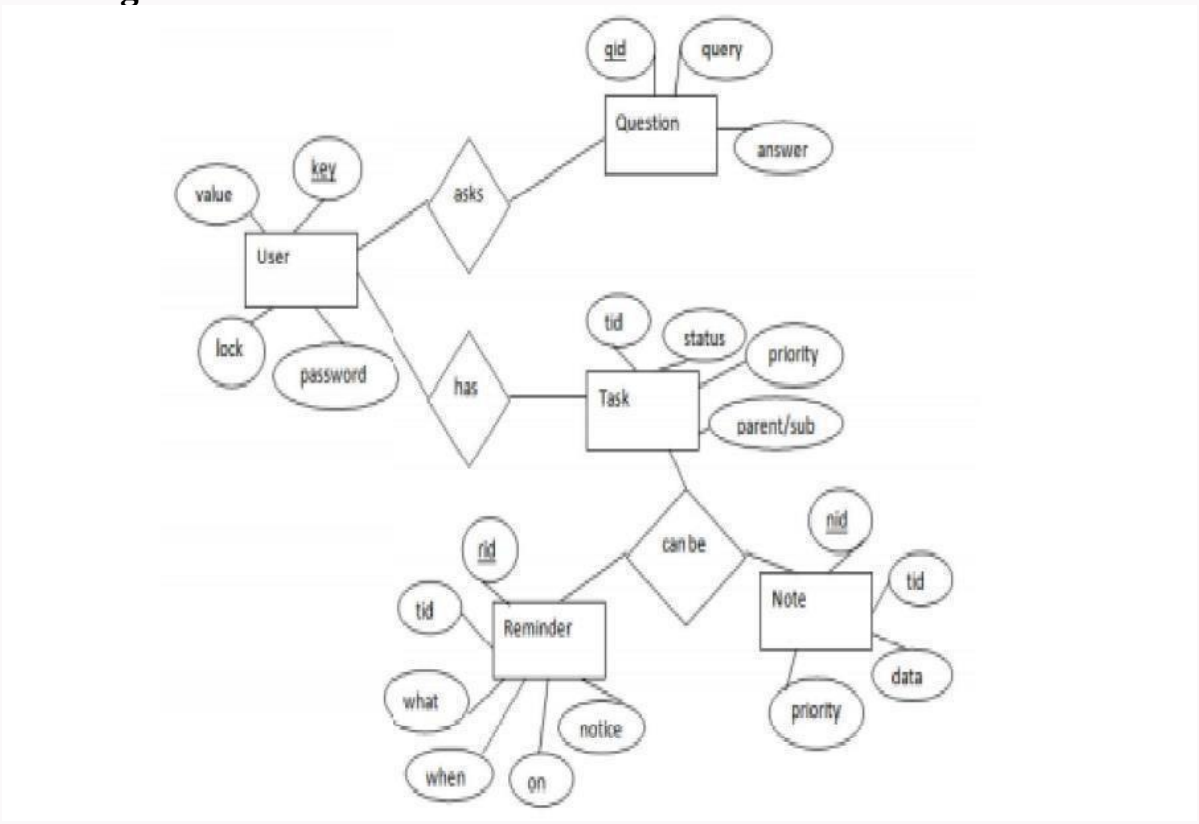
3.

Scope: This is an examination surveillance camera made using Python and AI. This is to be used during the examination to take the role of a virtual invigilator and monitor each and every moment of candidate; it emits a beep sound and also displays a message or notification if any candidate involves in unfair activity. This camera can be used in a variety of fields, including banking but here we are mainly focusing on examination point of view. A person can even utilize his or her own laptop as a security camera if he or she has the required code.

4. **Quality:** This system is come up with many qualities of advanced technology for users' interface. It recognizes serval gestures with the very less interval of time, means it's decisionmaking power for different gestures is smooth and strong enough. It is an ideal system so it does not need any kind of gloves, sensors, USB cables, markers but a human hand only with any kind of skin color.

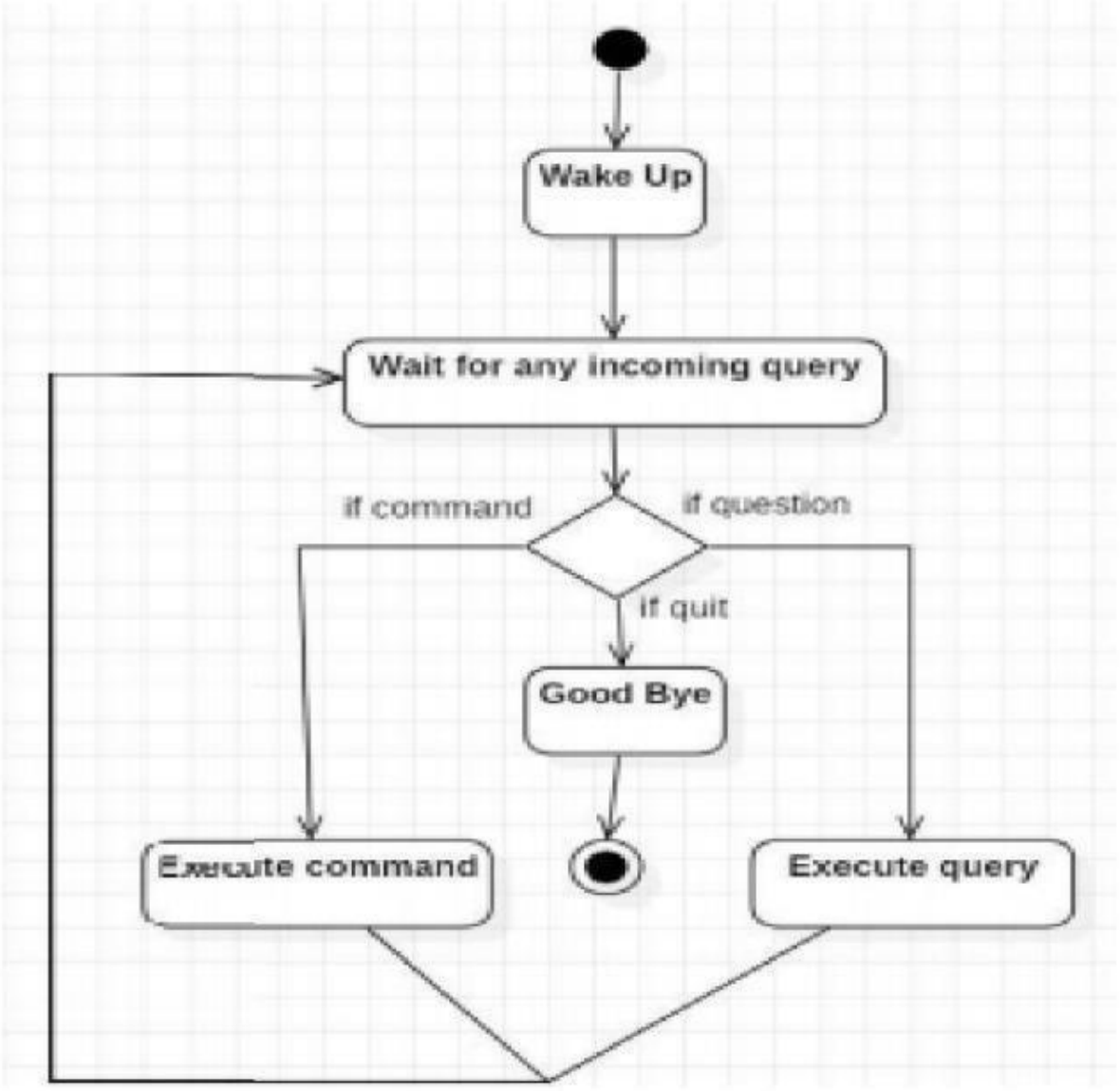
5. **Benefits:** Automating repeated tasks to a voice-activated personal assistant frees up the human time and resource s. Also, it can efficiently perform these mundane tasks with no errors, which often leads to an improvement in customer satisfaction. Whilevoice assistants are left to deal with routine tasks, humans can dedicate more time to duties where human intervention is required for successful business solutions and services.

ER Diagram



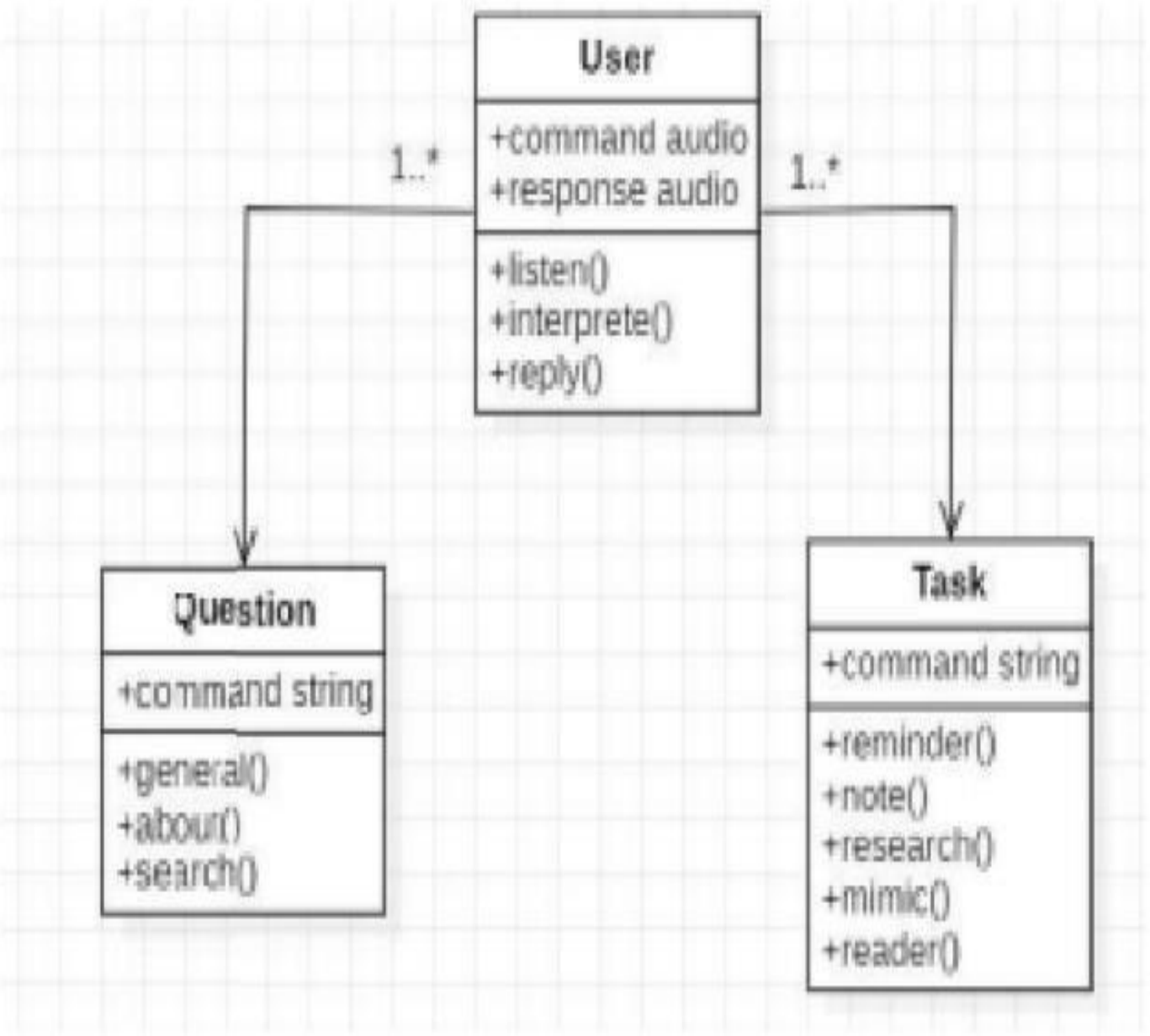
The above diagram shows entities and their relationship for a virtual assistant system. We have a user of a system who can have their keys and values. It can be used to store any information abou t the user. Sav. for kev “name” value can be “Jim”. For some kev’s user might like to keep secure . There he can enable lock and set a password (voice clip). Single user can ask multiple questions. Eac h question will be given ID to get recognized along with the query and its corresponding answer. Use r can also be having n number
of tasks. These should have their own unique id and status i.e. their current state. A task should also havea priority value and its category whether it is a parent task or child task of an older task.

Activity Diagram



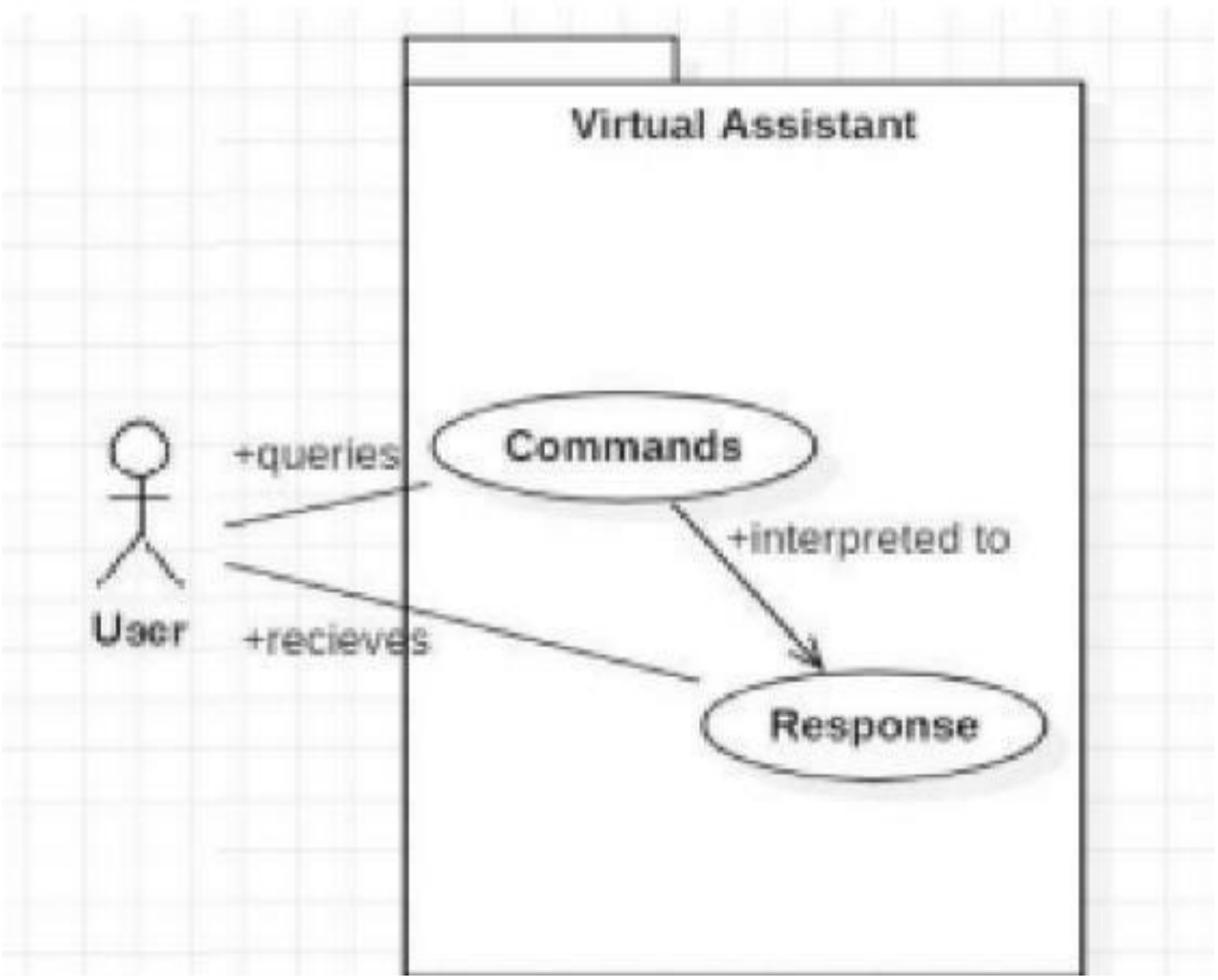
Initially, the system is in idle mode. As it receives any wake up cal it begins execution. The received command is identified whether it is a questionnaire or a task to be performed. Specific action is taken accordingly. After the Question is being answered or the task is being performed, the system waits for another command. This loop continues unless it receives quit command. At that moment, itgoes back to sleep.

Class Diagram



The class user has 2 attributes command that it sends in audio and the response it receives which isalso audio. It performs function to listen the user command. Interpret it and then reply or sends back response accordingly. Question class has the command in string form as it is interpreted by interpret class. It sends it to general or about or search function based on its identification. The taskclass also has interpreted command in string format. It has various functions like reminder, note, mimic, research and reader.

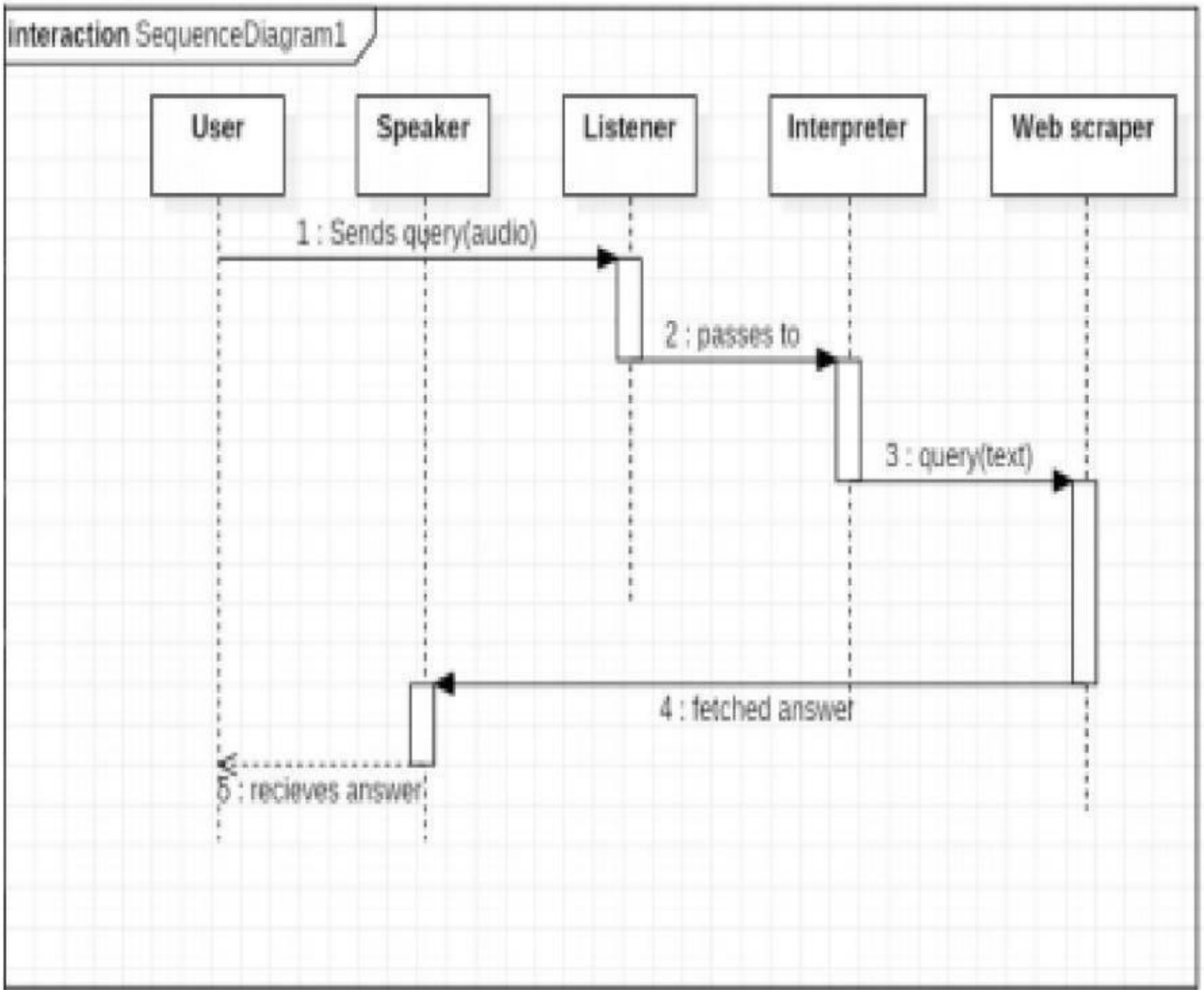
Use Case Diagram



In this project there is only one user. The user queries command to the system.
System theninterprets it and fetches answer. The response is sent back to the user.

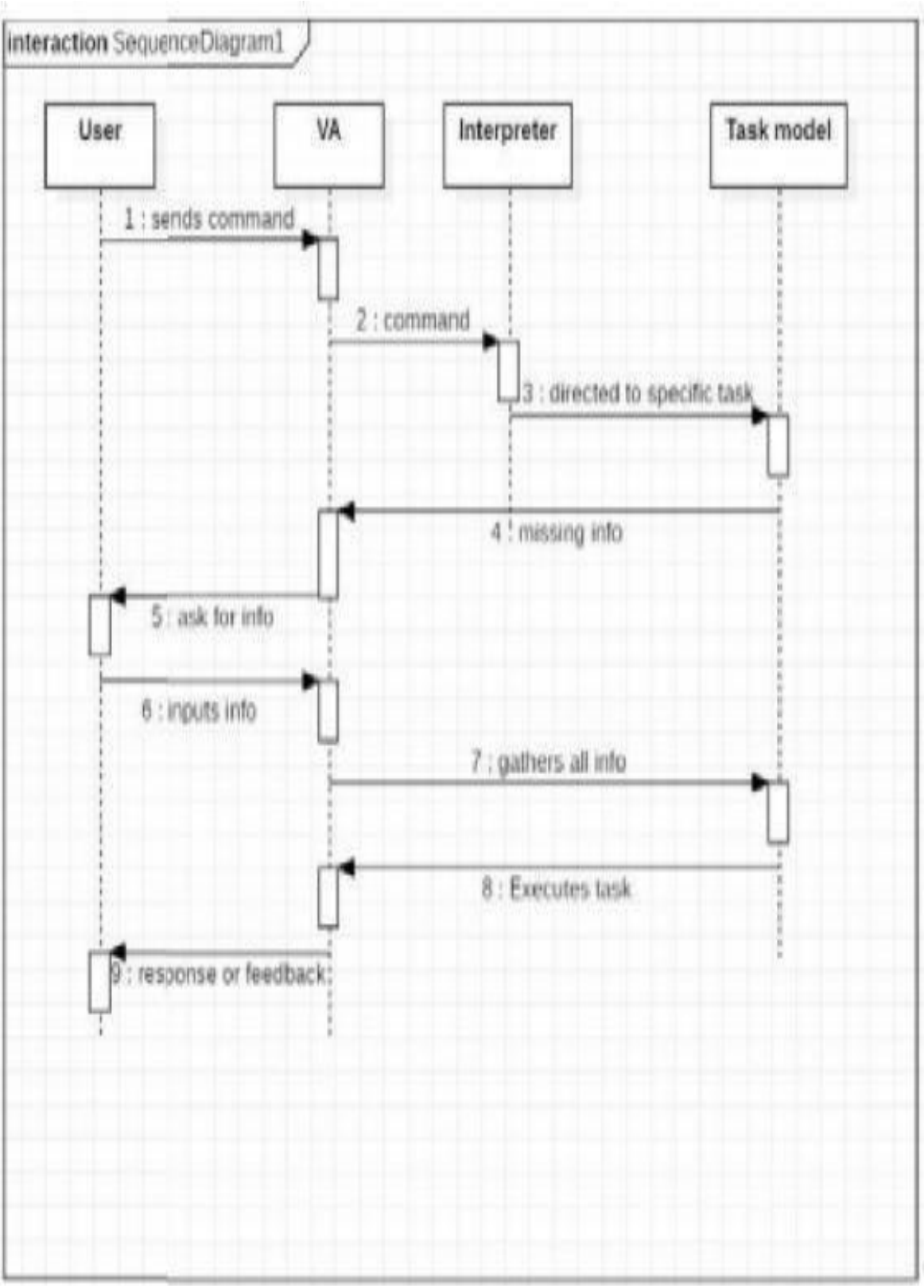
Sequence Diagram

A. Sequence diagram for Query-Response



The above sequence diagram shows how an answer asked by the user is being fetched from internet. The audio query is interpreted and sent to Web scraper. The web scraper searches and finds the answer. It is then sent back to speaker, where it speaks the answer to user.

B. Sequence diagram for Task Execution

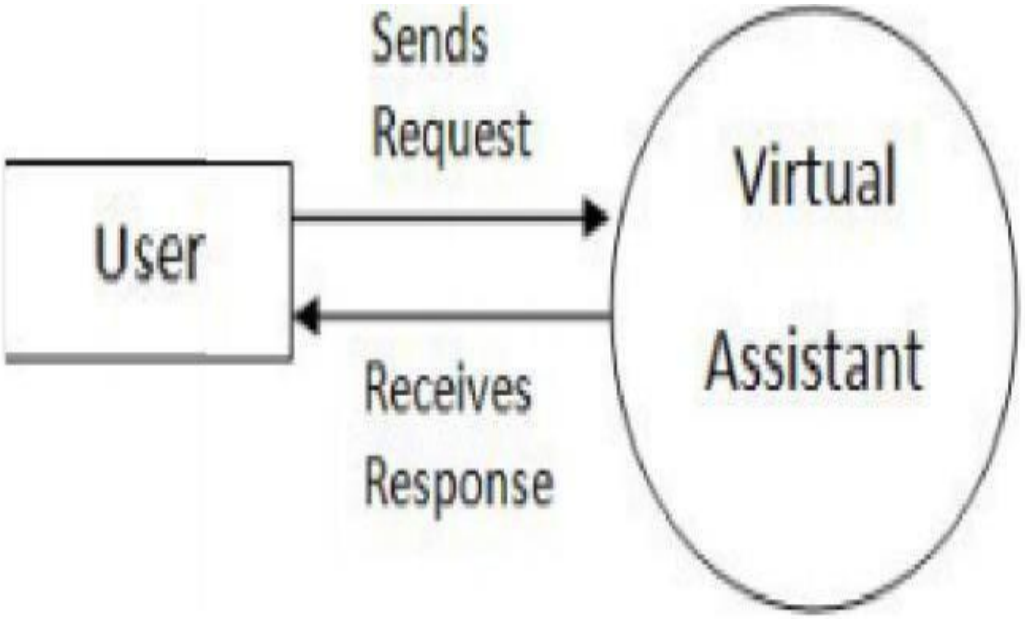


The user sends command to virtual assistant in audio form. The command is passed to the interpreter. It identifies what the user has asked and directs it to task executer. If the task ismissing some info, the virtual

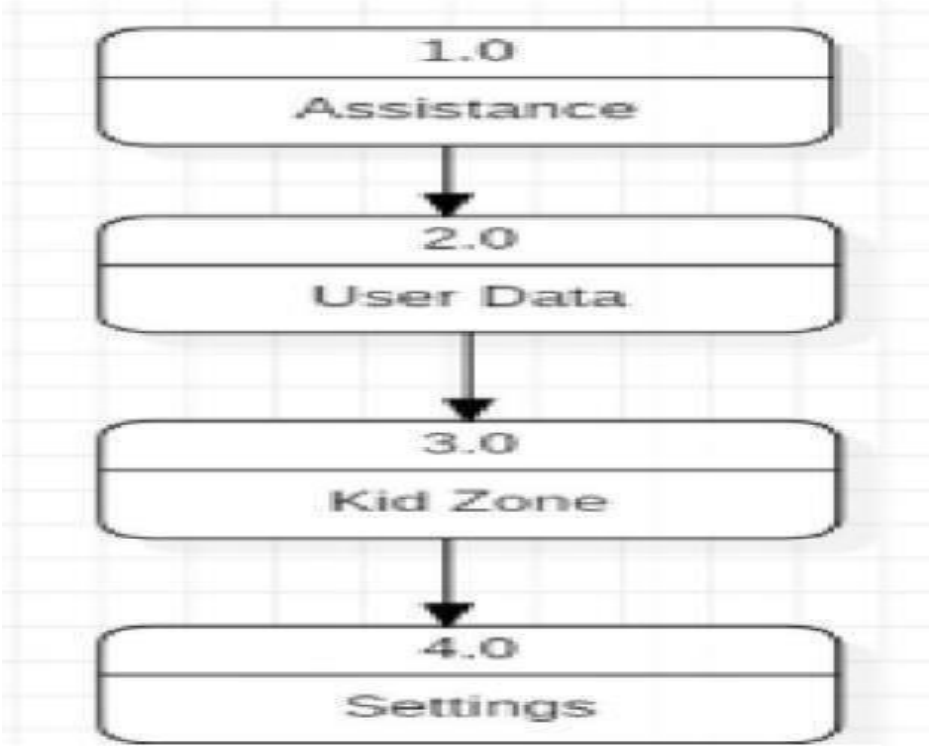
assistant asks user back about it. The received information is sent back to task and it is accomplished. After execution feedback is sent back to user.

Data Flow Diagram

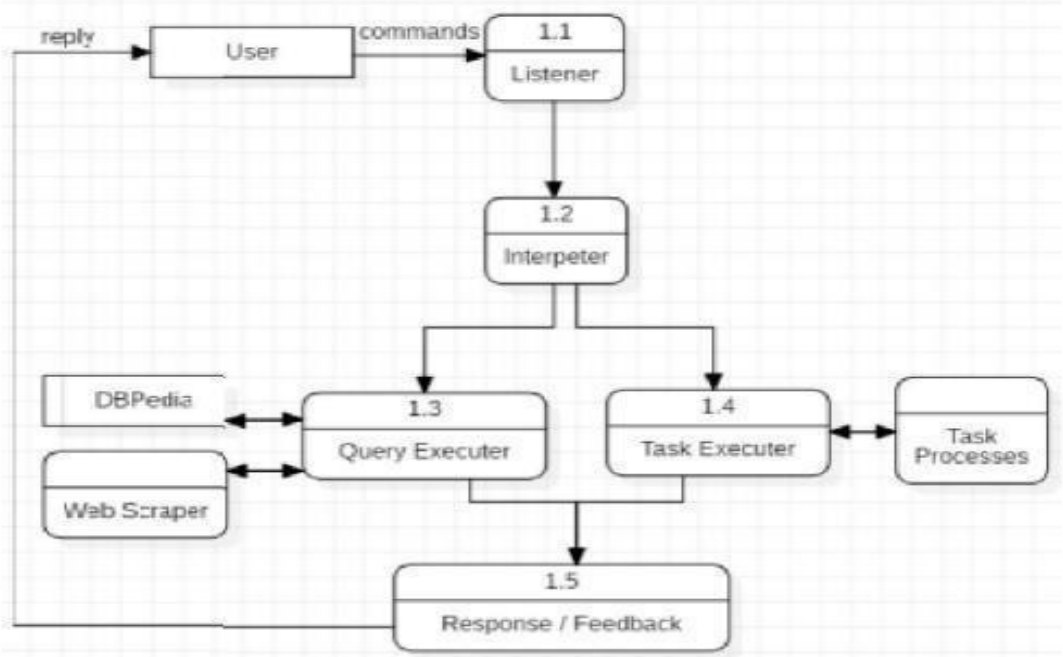
A. DFD Level 0 (Context Level Diagram)



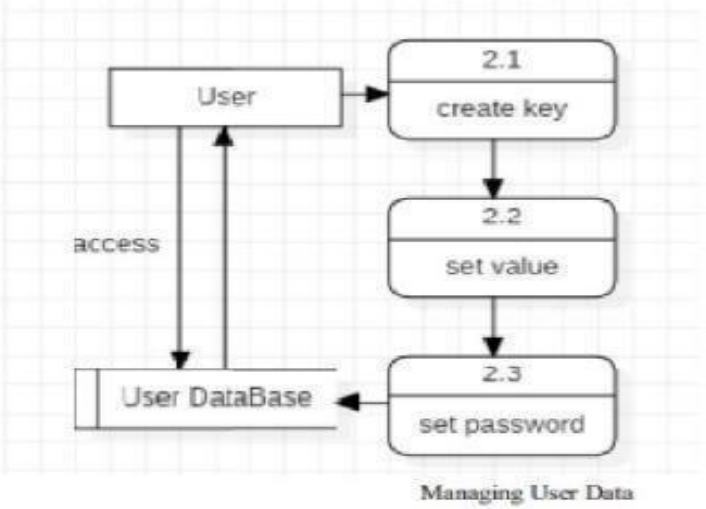
B. DFD Level 1



C. DFD Level 2

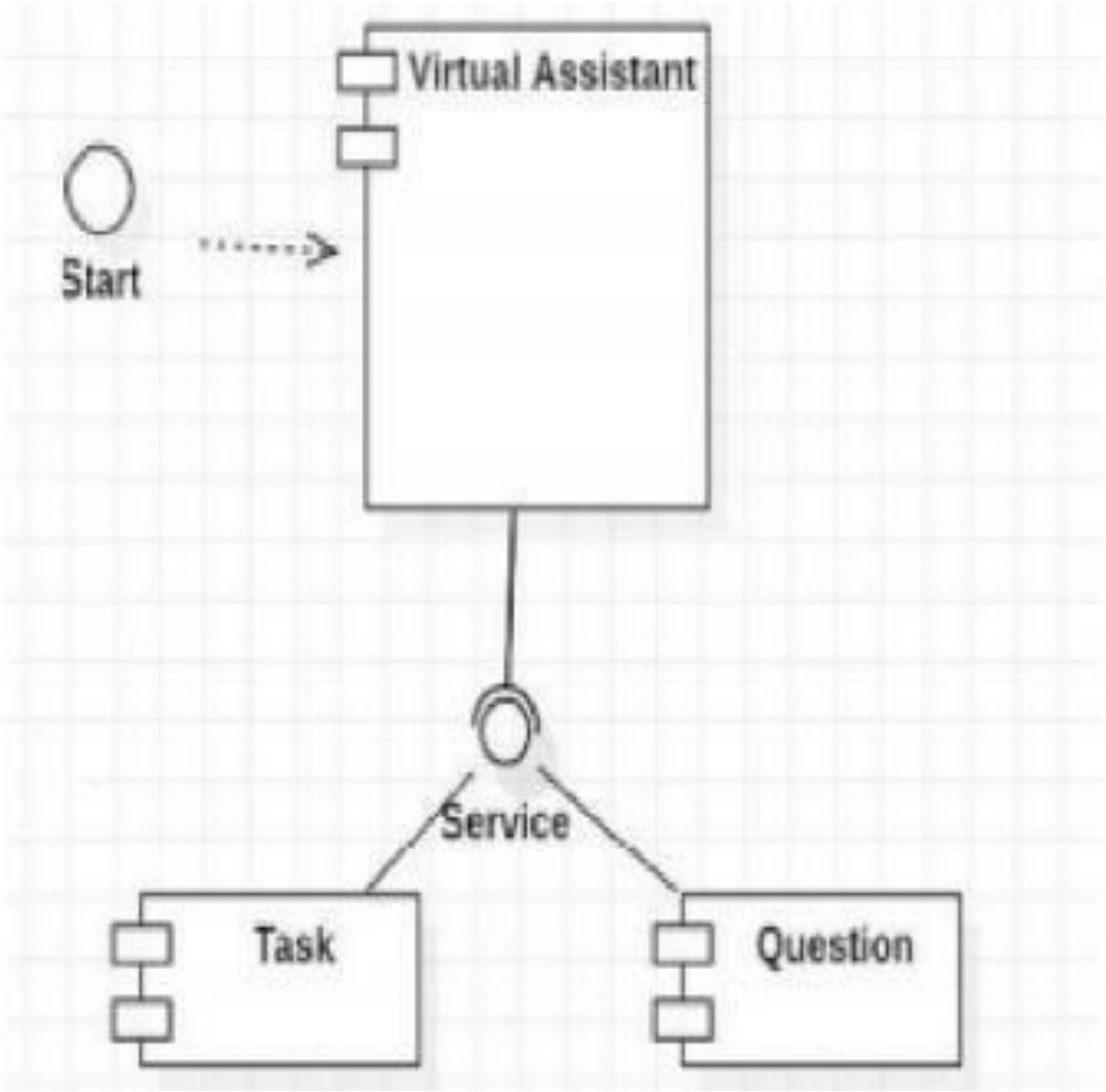


Data Flow in Assistance



Managing User Data

Component Diagram



The main component here is the Virtual Assistant. It provides two specific service,executing Task or Answering your question.

Results analysis and validation

Currently, the use of voice-assistants has been on the rise, but a user-centric usability evaluation of these devices is a must for ensuring their success. System Usability Scale (SUS) is one such popular usability instrument in a Graphical User Interface (GUI) scenario.

However, there are certain fundamental differences between GUI and voice-based systems, which makes it uncertain regarding the suitability of SUS in a voice scenario.

The present work has a twofold objective: to check the suitability of SUS for usability evaluation of voice-assistants and developing a subjective scale in line with SUS that considers the unique aspects of voice-based communication.

We call this scale as the Voice Usability Scale (VUS). For fulfilling the objectives, a subjective test is conducted with 62 participants. An Exploratory Factor Analysis suggests that SUS has a number of drawbacks for measuring the voice usability.

Moreover, in case of VUS, the most optimal factor structure identifies three main components: usability, affective, and recognizability and visibility.

The current findings should provide an initial starting point to form a useful theoretical and practical basis for subjective usability assessment of voice-based systems.

Given the uniqueness of voice-assistants and their difference from GUI-based systems, this work investigates two questions.

First, the suitability of SUS which is one of the most popular usability evaluation scales for GUI-based systems is checked for the voice-only context. Second, a new standardized scale is proposed keeping in mind the specific requirements of the voice-assistants.

SUS and the Usability of Voice-Assistants

SUS is a standard scale and an extremely popular tool for measuring the usability of GUI-based products. However, several new and different observations are obtained when using SUS for evaluating the usability of the voice-assistants.

First, when comparing SUS scores from previous studies with those currently obtained (Table, it is seen that the central tendencies of the distributions are not identical.

The mean SUS score reported in is 70.1 with a 99% confidence interval ranging from 68.7 to 71.5. However, presently, we obtain a mean SUS score of 63.69 with a 99% confidence interval ranging from 58.62 to 65.76.

Therefore, the confidence intervals for the two cases are non-overlapping, and the difference in mean is statistically significant ($p < 0.01$).

The second major difference is with respect to the factor structure of SUS. Extant research report that SUS is bi-dimensional having two components: usability (items 1, 2, 3, 5, 6, 7, 8, and 9), and learnability (items 4 and 10). This bi-dimensional nature of SUS has been found to be true in most of the testing scenarios .

However, for the current case, the results are substantially different. Instead on loading separately on a distinct component item 4 and 10 loads together with other items (2, 3, 6, and 8). Items 1, 5, 7, and 9 load together on the second component; however, the loadings are low for items 7 and 9 (less than 0.5).

Therefore, in case of the voice-assistants, the learnability component does not have any significance. The voice-only context provides a naturalistic and humanized environment when compared to the GUI systems that assist the users in completing their tasks.

The low loading of item 7 ("I imagine that most people would learn to use this voiceassistant very quickly") further indicates that the learnability dimension is of little importance. In fact, for the SUS dataset when the factor analysis is re-run eliminating items 7 and 9, better results are obtained. Thus, for voice-assistants, SUS is reduced to an 8-item scale.

Finally, the correlation observed between the SUS and ARS scales is too low. ARS scale was built from SUS to give it an adjective rating and make the original SUS scores meaningful.

Hence, historically SUS has shown to have a high correlation with ARS. Since ARS has just one item that measures the user-friendliness of a product, it is synonymous to the usability concept. However, the low correlation between the two scales is suggestive of the fact that SUS might not be a good measure of usability for the voice-assistants.

The VUS Scale

The factor analysis on the initial pool of items suggests three main components. These have been named as *Usability*, *Affective,* and *Recognizability & Visibility*. *Usability* dimension refers to the users' perceptions that the voice-assistants recognize them properly and do the tasks as instructed.

This is related to the voice-assistant's ability to correctly recognize what the users are speaking, correctly interpret the meaning of what the users are asking and act accordingly. In case of certain tasks that require a series of back-and-forth conversations between the user and the voice-assistant (for example during shopping or making a payment) for accomplishing the task, the users must know when to speak exactly.

In the absence of such a scenario, there will be a lack of synchronization, which will make the system difficult to use. Therefore, this dimension is not only related to how easily the voice- assistants and the users understand each other, but also being able to interact freely and easily that makes these systems easy to use.

This component accounts for the greatest proportion of the variance explained, indicating that it is one of the prime factors for evaluating the usability of voiceassistants. These days, voice-assistants are being used for a variety of purposes, both transactional and non-transactional. As such, the interactions must be clear, transparent and the information provided by these devices useful and timely.

The second dimension is *Affective*. We named it the affective dimension, since it portrays the satisfaction/frustration/expectation realization of the users after using the voice-assistants. This dimension explained the second-most proportion of variance in the factor analysis. The heuristic design principles for voice-assistants suggested by authors in also illustrate the importance of this factor.

The anthropomorphic features of voice can easily cause a disconfirmation in the users' perceptions of the capabilities of a voice-assistant versus its actual abilities. Moreover, it has been found that voice-based systems are typically sluggish than GUI's, because the interaction is through voice-only.

The naturalness and spontaneity of voice conversations increase the usability challenge for the voice-assistants as the users are accustomed to a certain way of communicating with other human beings and expect the same from these devices also. This makes the affective dimension highly relevant too for the purpose of usability evaluation.

The third and final factor is named as *Recognizability & Visibility*. Users must recognize the various functions and options provided by the voice-assistants just through interaction and affordance with the voice-assistants. The response given by the voice-assistants should be natural and easily understood by the users.

In this respect, the choice of appropriate vocabulary is important that can be understood by a variety of users. Previous work in also indicated problems with usability with respect to the accent of English spoken.

Since, the voice-assistants lack any type of a visual interface, it can lead to a higher cognitive load among the users as they must remember all the speech commands. This might make these systems difficult to customize based on the needs and preferences of the users.

The users may have to make many guesses while trying to customize the system using some trial-and-error method and might eventually abandon the task. Therefore, the visibility of the entire system might be affected.

Conclusion and future work

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.

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 neural

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.

The capabilities of voice assistants are continuously extending. Amazon and Google have provided platforms for developers in order to extend their assistants' capabilities. Similar to mobile apps, Amazon Skills and Google Actions, radically expand assistants' repertoire, allowing users to perform more actions with voiceactivated control.

According to Sheppard (2017), some key elements that distinguish voice assistants from ordinary programs are:

- NLP: the ability to understand and process human languages. It is important in order to fill the gap in communication between humans and machines

- The ability to use stored information and data and use it to draw new conclusions

- Machine learning: the ability to adapt to new things by identifying patterns Similarities and differences of devices and services regarding voice assistants have been studied in the literature (López et al., 2017; Këpuska and Bohouta, 2018).

In addition, as with any new revolutionary technology, scientific research and the educational community are considering whether these new devices can help the educational process. Something similar has happened before with personal computers and tablets (Algoufi, 2016; Gikas and Grant, 2013; Herrington and Herrington, 2007).

The purpose of our paper is to present findings regarding home usage of voice assistants and smart speakers, as well as some early attempts for using them for educational Voice Assistants and Smart Speakers in Everyday Life and in Education 475 purposes.

Although voice assistants are present in many homes, their use in school environments and for educational purposes is limited since there are many concerns regarding their privacy settings and data collection.

Study of home usage will provide insights regarding the ease of use of this new technology and how users perceive it. Furthermore, education can take place in formal or informal settings, thus it is evident to examine the use of voice assistants and smart speakers, inside or outside the classroom and by children, adults and elderly 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. 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.: NaturalLanguage 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/h igh-tech-gadgets/speech-recognition1.html

10. Nagesh Singh Chauhan,"Build Your First Voice Assistant",March,2019 https://towardsdatascience.com/build-your-firstvoiceassistant.

Abdolrahmani, A., Kuber, R., Branham, S. M. (2018). Siri Talks at You: An Empirical Investigation of VoiceActivated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 249–258). ACM. DOI: 10.1145/3234695.3236344 Algoufi, R. (2016).

Using Tablet on Education. World Journal of Education, 6(3), 113–119. Baldauf, M., Bösch, R., Frei, C., Hautle, F., Jenny, M. (2018). Exploring requirements and opportunities of conversational user interfaces for the cognitively impaired. In Proceedings of the 20th International Conference on human-computer interaction with mobile devices and services adjunct (pp. 119–126). ACM. DOI: 10.1145/3236112.3236128 Beirl, D., Yuill, N., & Rogers, Y. (2019).

Using Voice Assistant Skills in Family Life. In Lund, K., Niccolai, G. P., Lavoué, E., Gweon, C. H., & Baker, M. (Eds.), A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019,

Volume 1 (pp. 96–103). Lyon, France: International Society of the Learning Sciences. DOI: 10.22318/cscl2019.96 Bekmyrza, K. (2019). Using Chatbot to Increase Students' engagement in Education Process at High School. https://www.internauka.org/archive2/vestnik/11(61_2).pdf#page=68 Bunyard, S. (2019).

Assistance from Alexa: The social and material benefits of the Internet of Things. https:// scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1044&context=engl_176 Canalys. (2018). Media alert: Smart Speaker Installed Base to Hit 100 Million by End of 2018. https:// www.canalys.com/static/press_release/2018/090718 Media alert Smart

speaker installed base to hit 100 million by end of 2018.pdf Cho, E. (2019). Hey Google, Can I Ask You Something in Private? In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (p. 258).

ACM. Danovitch, J. H., Alzahabi, R. (2013). Children show selective trust in technological informants. Journal of Cognition and Development, 14(3), 499–513.

Davie, N., Hilber, T. (2018). Opportunities and Challenges of Using Amazon Echo in Education. International Association for Development of the Information Society. Dillon, T. (2018).

Creating an Alexa-Enabled Textbook Exercise: An Easy Approach to Custom Automatic Speech Recognition Application. KOTESOL Proceedings 2018, 259. Dousay, T. A., Hall, C. (2018).

Alexa, tell me about using a virtual assistant in the classroom. In EdMedia+ Innovate Learning (pp. 1413–1419). Association for the Advancement of Computing in Education (AACE). Druga, S., Williams, R., Breazeal, C., Resnick, M. (2017).

Hey Google is it OK if I eat you?: Initial explorations in child-agent interaction. In Proceedings of the 2017 Conference on Interaction Design and Children (pp. 595–600). ACM. DOI: 10.1145/3078072.3084330 Gikas, J., Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media.

The Internet and Higher Education, 19, 18–26. Griswold, A. (2018). Even Amazon is surprised by how much people love Alexa. (February 2018). https:// qz.com/1197615/even-amazon-is-surprised-by-how- much-peoplelove-alexa/ Herrington, A., Herrington, J. (2007).

Authentic mobile learning in higher education. In Proceedings of the AARE 2007 International Educational Research Conference. Fremantle, Western Australia, 1–9. DOI: 10.1109/ICNICONSMCL.2006.103