JARVIS: A PC Voice Assistant

Ijasrw editor

International Journal of Advance Study and Research Work

JARVIS: A PC Voice Assistant

Jash Vora[1], Deepak Yadav[2], Ronak Jain[3] & Jaya Gupta[4]
Department of Computer Engineering, A. P. Shah Institute of Technology, G.B.Road, Kasarvadavli, Thane (W),
Mumbai-400615, University of Mumbai.
Email Id: jashvora100@gmail.com, ydeepak141097@gmail.com, jronak083@gmail.com & jdgupta@apsit.edu.in

DOI: 10.5281/zenodo.5323068

Abstract

The project aims to develop a private assistant for Computer. PC Personal Assistant draws its inspiration from virtual assistants like Google Assistant for Android, Siri for iOS. It provides a user-friendly interface for completing a spread of tasks by employing certain well-defined commands. Users can interact with the assistant through voice commands. As a private assistant, it assists the end-user with day-to-day activities like a general human conversation, searching queries, reading the latest news, translating words, live weather, sending mail through voice. The software uses a device's microphone to receive voice requests while the output takes place at the system's speaker. It's a mixture of various technologies: voice recognition, voice analysis, and language processing. PC Personal assistant is built mainly using python.

Key Words: Speech Recognition, Wikipedia, Speech Synthesis, Wolfram Alpha, Voice Search.

Introduction

Our digital life is decided by innovations. Especially in recent years, more innovative technologies were developed to ease our professional lifestyle. Intelligent Personal Assistant is proved to be the most vital innovation in terms of easing our lives and providing a hands-free experience. We are building a PC Personal Assistant that works on voice commands and executesthe user query [2]. Our project, PC Personal Assistant is built mainly using python. The software uses a device's microphone to accept voice requests while the output takes place at the system's speaker. It's a mixture of varied technologies: voice recognition, voice analysis, and language processing. When a user asks an assistant to perform a task, the natural language audio signal is converted into digital data which will be analyzed by the software [8]. Once digitized, several models are used to transcribe the audio signal to textual data. Keywords are used to perform certain tasks and if those keywords are present within the text then the particular task is performed. For instance, the 'translate' keyword is employed for the translation of the word from one language to a different so if user audio contains that keyword then the translation function is going to be activated and translation tasks are going to be performed by an assistant. So keywords are defined for particular tasks and if that word is present in your audio the actual tasks are going to be performed. Tasks that our Pc assistant can perform are searching the information from Wikipedia and reading the information, fetching top news from Times of India and reading the news, telling the present weather of a particular city, translating a word or sentence into the desired language, computational task, telling you the present time, capturing a photo from your camera, etc. Various APIs are wont to perform the task. For instance, Wolfram Alpha API is employed to perform computational tasks. Python also provides various libraries for our help. Speech Recognition library is employed to perform speech to text conversion, Wikipedia library is employed to urge information from Wikipedia, pyttsx3 library is employed to perform the text to speech, etc [8]. All the tasks are within the textual form which is then converted into an audio signal. A Text-to-speech Engine converts the text into phonemic representation, and then it converts the phonemic representation to waveforms which will be output. This is the overall workflow of our project [3].

Literature Review

A. Speech Synthesis

Speech synthesis may be a technique of artificial production of human speech. Speech synthesis produces audio stream as output. A speech recognizer on the opposite hand does the opposite work. It takes an audio stream as input then turns that audio stream into text transcription. The computer software that's used for speech synthesis is named a speech synthesizer, and it is often implemented during software or hardware products. Synthesized speech is often created by combining the pieces of recorded speech that are stored in a database. A Text-to-speech system transforms normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech as shown inFigure1.

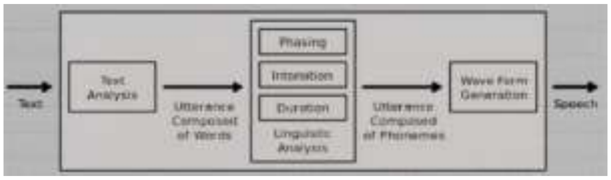


Figure 1: Working of Text-to-Speech System.

The quality of speech synthesizer is judged by the similarity of the human voice and by its ability to be understood clearly by the user [2]. The foremost important qualities of a speech synthesizer are naturalness and intelligibility. Naturalness defines how closely the output looks like human speech, while intelligibility is that the ease with which the output is understood. The perfect speech synthesizer has naturalness also as intelligibility. Speech synthesis systems try to maximize both characteristics.

B. Synthesizer Technologies

The two fundamental technologies that generate synthetic speech waveform are concatenative synthesis and formant synthesis. Each technology hasits pros and cons, and therefore the intended uses of a synthesis system will typically determine which approach is employed. Some popular speech recognition software is Siri, Cortana, Google Now, etc.

Proposed System

Based on the study of this system, the proposed systemaims to simplify basic operations for the user, users with faulty hardware, users who could be too busy to perform certain operations themselves, elderly people, and even users with sight or motor disabilities.

For example, a teacher could be scoring exam papers and remembers that he must book a flight, rather than leaving the work he's doing he could simply tell the Voice Assistant application to assist him by saying, “find nearby airports” and therefore the application willhelp the user open his browser and find airports on the brink of him and other flight details, inherentlynullifying the necessity for him to do it himself. the appliance also will allow him (in best-case scenarios) toperform this task significantly faster than he would havedone otherwise. The appliance also possesses speech synthesizing capabilities to offer the user the impressionthat he's talking and dealing with an actual assistant.
Speech Recognizer will take an input file as a voice andmap it into its textual representation. This feature is specially designed for Blind people [3]. The native userwho barely knows to work a laptop can easily open this application using voice commands. It responds to basic commands like Open Applications, Close Applications,Connect Google Send Mail to the respective person. Allthis will be performed on the voice commands of the topuser without internet connectivity [3].

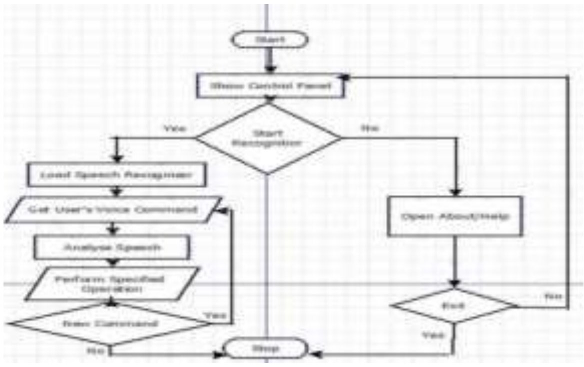


Figure 2: Flow Chart of Project

Implementation

A. Speech Recognition

The first component of the speech recognition system is speech, speech needs to be converted from its physical form to an electrical signal with the help of a microphone which is then digitized using several models [1].

Most modern speech recognition systems rely on Hidden Markov Model (HMM). HMM, the approach works on the idea that a speech signal, when viewed on a brief timescale (say, ten milliseconds), is often reasonably approximated as a stationary process that is, a process during which statistical properties don't change overtime.

In a Hidden Markov Model, the speech signal is split into 10-millisecond fragments. The spectrum of every fragment in HMM, which is a plot of the signal's power function of frequency, and is mapped to a vector of real numbers referred to as cepstral coefficients [1]. The dimension of this vector is typically small sometimes it isas low as 10, although more accurate systems may have dimension 32 or more. the ultimate output of the Hidden Markov Model may be a sequence of those vectors.

To decode the speech into text, vectors are grouped and matched to at least one or more phonemes which may bea fundamental unit of speech. This calculation requires training, since the sound of a phoneme varies fromperson to person, and even varies from one utterance to a different by an equivalent speaker. A special algorithmis applied to work out the foremost likely word (or words) that produce the given sequence of phonemes. In Python, we have a Speech Recognition library that performs Speech-to-text conversion. the primary step in Speech Recognition is to acknowledge the speech. The Recognizer Class in Speech Recognition library is employed to acknowledge the speech. As discussed earlier we'd like a microphone to convert physical sound into an electrical signal so for that we've Microphone Class in the Speech Recognition library. To capture the input from the microphone we'd like to use another method called 'listen ()'.Finally, we recognize the microphone input.

B. Regular Expression

A regular expression may be a sequence of characters that helps us in searching a pattern in a string. Each character during a regular expression is either a metacharacter, having special meaning, or a daily character that features a literal meaning [3].

For instance, within the regular expression b., 'b' could also be a literal character that matches just 'b', while '.' could also be a metacharacter that matches every character except a newline. Together, metacharacters and literal characters are often wont to identify the text of agiven pattern or process a variety of instances of it. The listof metacharacter include:. ^ $ * + ? { } [] \ | (). A regex processor translates a daily expression into anindoor representation which will be executed and matchagainst a string that represents the text that is being searched in [4]. we've used Regular expression in our project to extract YouTube links.

In Python, there's a library called 're' which is used in Regular expression.

1. findall() - it's wont to look for "all" occurrences that match a given pattern. it'll check all the lines of an input string.
2. match() - Determine if the regular expression matchesat the start of the string.
3. search() - Scan through a string, trying to find any location where RE matches
4. group() - Returns the string matched by the RE
5. start() - Return the starting position of the match
6. end() - Return the ending position of the match
7. span() - Returns a tuple containing the (start, end) positions of the matched string.

C. Urllib.parse

Urllib.parse defines a standard interface to break Uniform Resource Locator (URL) strings up into components (addressing scheme, network location, path, parameters, query, and fragment), to combine the components back into a URL. We have used urllib.parse to retrieve information from Wolfram alpha API and to get user query-based youtube URL.

1. urllib.parse.urlparse() - function is used to split url string into its components, or on combining URL components into URL string.
2. urllib.parse.parse_qs() - function is used to parse aquery string given as a string argument. Data are returnedas a dictionary.

D. Urllib.parse

Urllib.parse defines a typical interface to interrupt Uniform Resource Locator (URL) strings up into components (addressing scheme, network location, path, parameters, query, and fragment), to combine the components back to a URL. We have used urllib.parse toretrieve information from Wolfram alpha API and to getuser query-based youtube URL.

3. urllib.parse.urlparse() - function is used to split url string into its components, or on combining URL components into URL

string.

4. urllib.parse.parse_qs() - function is used to parse a query string given as a string argument. Data are returnedas a dictionary.
5. urllib.parse.urlencode() - this function converts a mapping object, which may contain str objects, to apercent-encoded ASCII text string.

E. Translation

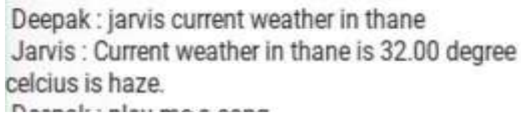
A translation system is a neural network that takes a sentence as input and produces a translation of thatsentence as output. The first neural network learns how to encode the sequence of words that represents meaninginto an array. The second neural network learns how to decode those numbers back into a sequence of words thatmean the same thing [12]. The encoder takes in words in one language and the decoder outputs words into different languages. So in effect, the model learns to map from one human language to a special language via intermediate numerical encoding.

Result

The PC Based Voice Assistant will empower the client to:

1. Weather-related Information

- If the user wants to know weather information of aparticular city or his/her home city this feature comes in handy. Weather-related information of anycity can be provided with the help of this feature.
- We have used regular expressions to extract what the user said. To initiate this function user has to say current weather or weather in XYZ city.
- re. group() function will extract the city name from the sentence and pass it to the weather API and the user can get Weather-related information.

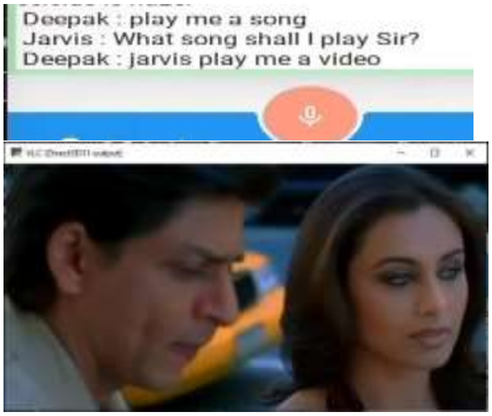


```
Deepak : jarvis current weather in thane
Jarvis : Current weather in thane is 32.00 degree
celcius is haze.
```

Fig 3: Output screen of Showing Weather-related Information.

2. Play songs and videos from Youtube

- We always thought that if we could listen to ourfavorite music by just providing its name through our voice command.
- We have provided this functionality in our project where users can listen to their favorite music just by providing the name of the song bytheir voice command and it will be played in thebackground.
- The name of the function in our Jarvis App is “youtube_Music()”. The control of the programwill be passed to this function when the user will say ‘play me a song.’
- This function will ask the user which song youwant to play and by providing that the user canlisten to that particular song.
- We have used web parsing and regular expression to achieve this. Users can also control the music through his voice. For example, if a user wants to increase the volume or decrease the volume or he/she wants to pause or resume the song all this they can do using voice-command.
- Along with youtube music if the user wants to play videos of his/her choice from youtube into his/her system’s VLC Media Player they can dothat with the help of our “youtube_Video()” functionality.
- Similar to youtube music user has to say ‘play me a video’ and youtube_Video gets control of the program and gets executed. All the functionality like play, pause, volume increase, volume decrease is there in video play as well.

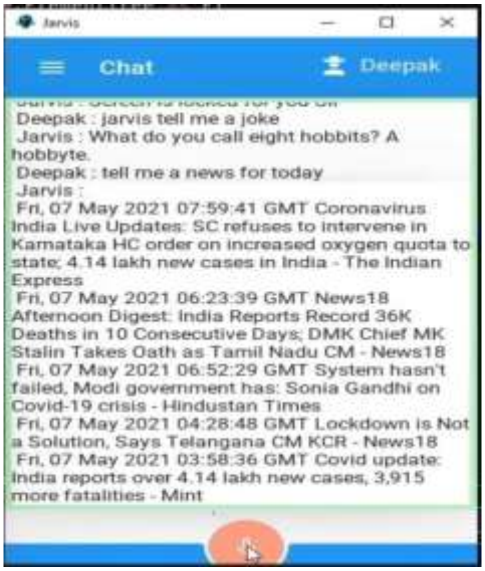


```
Deepak : play me a song
Jarvis : What song shall I play Sir?
Deepak : jarvis play me a video
```

Fig 4: Output screen of Playing song andvideo.

3. **News For The Day**

- If the user wants to keep him/her up-to-date byreading news he/she can do that with just one command i.e 'news for today' and the top 5 news of the day will be available to the user.
- Python's Beautiful Library is used to convertXML data into the user-readable format

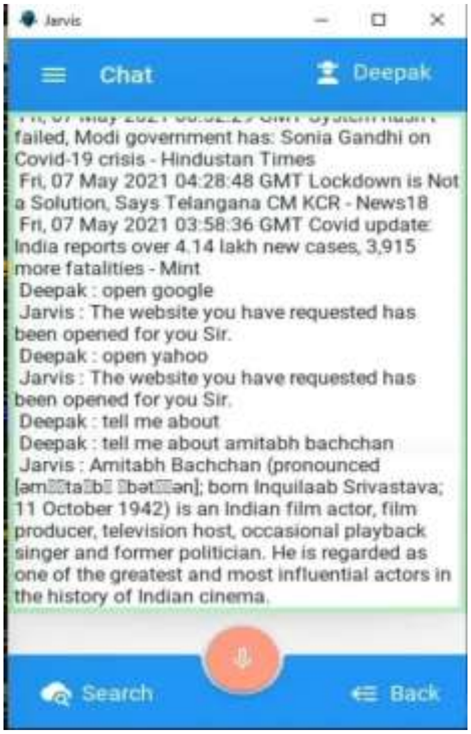


```
 Deepak : jarvis tell me a joke
 Jarvis : What do you call eight hobbits? A
hobbyte.
 Deepak : tell me a news for today
 Jarvis :
 Fri, 07 May 2021 07:59:41 GMT Coronavirus
India Live Updates: SC refuses to intervene in
Karnataka HC order on increased oxygen quota to
state; 4.14 lakh new cases in India - The Indian
Express
 Fri, 07 May 2021 06:23:39 GMT News18
Afternoon Digest: India Reports Record 36K
Deaths in 10 Consecutive Days; DMK Chief MK
Stalin Takes Oath as Tamil Nadu CM - News18
 Fri, 07 May 2021 06:52:29 GMT System hasn't
failed, Modi government has: Sonia Gandhi on
Covid-19 crisis - Hindustan Times
 Fri, 07 May 2021 04:28:48 GMT Lockdown is Not
a Solution, Says Telangana CM KCR - News18
 Fri, 07 May 2021 03:58:36 GMT Covid update:
India reports over 4.14 lakh new cases, 3,915
more fatalities - Mint
```

Fig 5: Output screen of Showing CurrentNews.

4. **Wikipedia Search**

- Wiki Search on any topic is now very easy using our app's voice to search functionality.
- One can Wiki search on any topic of his/herchoice using his/her voice command.
- Our Jarvis App will read out the content forthe user.

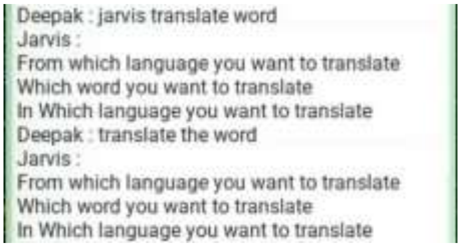


```
failed, Modi government has: Sonia Gandhi on
Covid-19 crisis - Hindustan Times
 Fri, 07 May 2021 04:28:48 GMT Lockdown is Not
a Solution, Says Telangana CM KCR - News18
 Fri, 07 May 2021 03:58:36 GMT Covid update:
India reports over 4.14 lakh new cases, 3,915
more fatalities - Mint
 Deepak : open google
 Jarvis : The website you have requested has
been opened for you Sir.
 Deepak : open yahoo
 Jarvis : The website you have requested has
been opened for you Sir.
 Deepak : tell me about
 Deepak : tell me about amitabh bachchan
 Jarvis : Amitabh Bachchan (pronounced
[əm▯▯ta▯b▯ ▯bət▯▯ən]; born Inquilaab Srivastava;
11 October 1942) is an Indian film actor, film
producer, television host, occasional playback
singer and former politician. He is regarded as
one of the greatest and most influential actors in
the history of Indian cinema.
```

Fig 6: Output screen of performingWikipedia Search.

5. Translate a word

- Users can translate a word/sentence from one language to another by providing the -translate and language in which he/she wants totranslate.
- Users can do language translation by just saying'translate a word' and the control of the the program will be transferred to the "translator_Word()" function andthe word/sentence will be converted to user-specific language.



```
Deepak : jarvis translate word
Jarvis :
From which language you want to translate
Which word you want to translate
In Which language you want to translate
Deepak : translate the word
Jarvis :
From which language you want to translate
Which word you want to translate
In Which language you want to translate
```

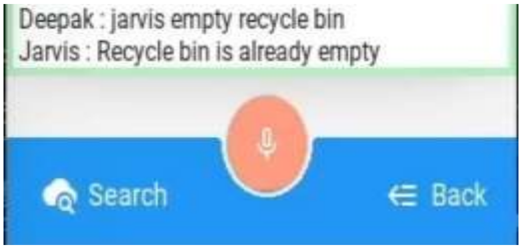
Fig 7: Output screen of Translating a word.

6. Take a photo and Screenshot

- Users can take his/her photo by saying 'take a photo' or 'camera' and his/her camera will be open.
- Users can take a screenshot of the screen by justsaying 'take screenshot' or 'screenshot'.
- The Screenshot is saved in the Pictures folderby default.

7. Empty Recycle Bin

- Users can empty his/her computer recycle bin by saying “empty recycle bin” and it will be done for the user.
- Python’s ‘winshell’ library is used to empty recycle bin.



Deepak : jarvis empty recycle bin
Jarvis : Recycle bin is already empty

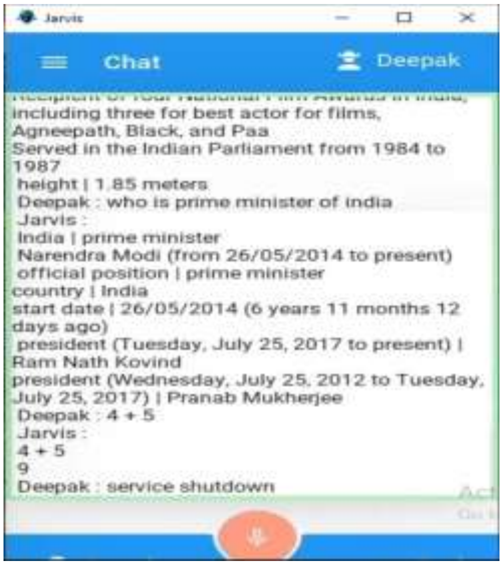
Fig 8: Output of Empty Recycle BinCommand.

8. To-Do List

- Users can create his/her to-do list using our Jarvis Voice Assistant and add things that he/she wants the Voice Assistant to remind him/her. Users can say “stop” or “done” or “that’s it” to stop adding elements to the list. Once added if the user wants to add to the existing list he/she can do that or they can also create a new list.
- Jarvis will remind the users about their list when the user will say ‘remind the list’. If there are multiple lists created by the user Jarvis will ask the user which list he/she wants to see.

9. Computational Query

- If a user wants to solve or know about ‘Mathematics’ or ‘Science and Technology’ or ‘Society and Culture’ or ‘Everyday Life’ he/shecan easily get that done by just saying ‘who is XYZ’ or ‘how…’ or ‘what…’
- We have used the ‘Wolfram Alpha’ API to extractthe content that wolfram alpha has.



```
including three for best actor for films,
Agneepath, Black, and Paa
Served in the Indian Parliament from 1984 to
1987
 height | 1.85 meters
 Deepak : who is prime minister of india
 Jarvis :
 India | prime minister
 Narendra Modi (from 26/05/2014 to present)
 official position | prime minister
country | India
start date | 26/05/2014 (6 years 11 months 12
days ago)
 president (Tuesday, July 25, 2017 to present) |
Ram Nath Kovind
president (Wednesday, July 25, 2012 to Tuesday,
July 25, 2017) | Pranab Mukherjee
 Deepak : 4 + 5
 Jarvis :
 4 + 5
 9
 Deepak : service shutdown
```

Fig 9: Wolfram alpha’s Result Output.

Conclusion

Voice assistants are useful in many fields such as education, daily life application, home appliances, etc. and the voice assistant is also useful for illiterate people they can get any information just by saying to the assistant, luxury is available for people, thanks to

AI-based voice assistants.

Through this voice assistant, we have automated various services using a single line command. It eases most of the tasks of the user like searching the web, retrieving weather forecast details, translating words from one language to another language, accessing youtube videos, sending mail through voice, and solving computational queries. We aim to make this project a complete User Interface based project and give the user all its queries on the very same User Interface.

With the advancements in technology, particularly in Artificial Intelligence, we can extend the scope of the project with Home Automation.

References

[1]. V. Mitra, H. Franco, M. Graciarena and D. Vergyri, "Medium-duration modulationcepstral feature for robust speech recognition," *2014 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, 2014, pp. 1749-1753, doi: 10.1109/ICASSP.2014.6853898.

[2]. A. Acero, "An overview of text-to-speech synthesis," *2000 IEEE Workshop on Speech Coding. Proceedings. Meeting the Challenges of the New Millennium (Cat. No.00EX421)*,Delavan, WI, USA, 2000, pp. 1-, doi: 10.1109/SCFT.2000.878372.

[3]. T. Kim, "Short Research on Voice Control System Based on Artificial Intelligence Assistant," *2020 International Conference on Electronics, Information, and Communication(ICEIC)*, Barcelona, Spain, 2020, pp. 1-2, doi:10.1109/ICEIC49074.2020.9051160.

[4]. Jin He, Rui Yu, Xinsheng Wang and Lina Huang, "Validation of query expression based on Regular Expression," 2011 International Conference on Computer Science and ServiceSystem (CSSS), Nanjing, China, 2011, pp. 1879-1882, doi: 10.1109/CSSS.2011.5974145.

[5]. I. Budiselic, S. Srbljic, and M. Popovic, "RegExpert: A Tool for Visualization of Regular Expressions," EUROCON 2007 - The International Conference on "Computer as a Tool", Warsaw, Poland, 2007, pp. 23872389, doi: 10.1109/EURCON.2007.4400374.

[6]. D. M. Thomas and S. Mathur, "Data Analysis by Web Scraping using Python," 2019 3rd International Conference on Electronics, Communication, and Aerospace Technology(ICECA), Coimbatore, India, 2019, pp. 450- 454, doi: 10.1109/ICECA.2019.8822022.

[7]. S. S. Chowdhury, A. Talukdar, A. Mahmud and T. Rahman, "Domain-Specific Intelligent Personal Assistant with Bilingual Voice Command Processing," *TENCON 2018 - 2018 IEEE Region 10 Conference*, Jeju, Korea (South), 2018, pp. 0731-0734, doi:10.1109/TENCON.2018.8650203.

[8]. K. N., R. V., S. S. S., and D. R., "Intelligent Personal Assistant - Implementing Voice Commands enabling Speech Recognition," *2020 International Conference on System, Computation, Automation, and Networking (ICSCAN)*, Pondicherry, India, 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262279.

[9]. G. A. Triantafyllidis and M. G. Strintzis, "A least-squares algorithm for efficient context-based adaptive arithmetic coding," *ISCAS 2001.The 2001 IEEE International Symposium onCircuits and Systems (Cat. No.01CH37196,*Sydney, NSW,Australia,2001, pp.169172vol2,doi:10.1109/ISCAS.2001.92 1034.

[10]. Saadman Shahid Chowdury, Atiar Talukdar, Ashik Mahmud, Tanzilur Rahman³Domain specific Intelligent personal assistant with bilingual voice command processing IEEE2018.

[11]. Polyakov EV, Mazhanov MS, AY Voskov, LS Kachalova MV, Polyakov SV ³Investigation anddevelopment of the intelligent voice assistant forthe IOT using machine learning Moscow workshop on electronic technologies, 2018.

[12]. Khawir Mahmood, Tausfer Rana, Abdur Rehman Raza³Singular adaptive multi-role intelligent personal assistant (SAM-IPA) for human-computer interaction´ Internationalconference on open source system and technologies, 2018.

[13]. R. Sangpal, T. Gawand, S. Vaykar, and N.Madhavi, "JARVIS: An interpretation of AIMLwith the integration of gTTS and Python," 2019 2nd International Conference onIntelligent Computing, Instrumentation and Control Technologies(ICICICT),Kannur, India,2019,pp. 486489, doi:10.1109/ICICICT46008.2019.8993344.

[14]. P. Bose, A. Malpthak, U. Bansal and A. Harsola,"Digital assistant for the blind," 2017 2nd International Conference for Convergence in Technology(I2CT), Mumbai, 2017, pp. 1250- 1253, doi: 10.1109/I2CT.2017.8226327.

[15]. M. R. Sultan, M. M. Hoque, F. U. Heeya, I. Ahmed, M. R. Ferdouse and S. M. A. Mubin,"Adrisya Sahayak: A Bangla Virtual Assistant for Visually Impaired," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), DHAKA,Bangladesh,2021, pp. 597-602,doi: 10.1109/ICREST51555.2021.9331080.

[16]. S. Kumari, Z. Naikwadi, A. Akole, and P. Darshankar, "Enhancing College Chat Bot Assistant with the Help of Richer Human- Computer Interaction and Speech Recognition," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp.427433, doi:10.1109/ICESC48915.202 0.9155951.