2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) | 978-0-7381-3042-2/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICREST51555.2021.9331080

Adrisya Sahayak: A Bangla Virtual Assistant for Visually Impaired

Md.Rakibuz Sultan[1], Md Moinul Hoque[2], Farah Ulfath Heeya[3], Iftiquar Ahmed[4], Md.Redwanul Ferdouse[5], Shikder Mejbah Ahmed Mubin[6]
[1-5]*Department of Computer Science and Engineering,*
[6]*Department of Electrical and Electronic Engineering,*
Ahsanullah University of Science and Technology,
Dhaka, Bangladesh
[1]pranto055@gmail.com, [2]moinul@aust.edu, [3]farah.heeya@gmail.com, [4]oveaust33@gmail.com,
[5]redwanhriday1@gmail.com, [6]mmejbahahmed@gmail.com

***Abstract*—Visually impaired persons have to encounter numerous complications in their mundane activities for imperfect vision. Though there are different technologies, Systems interaction procedure or expenditure is the key factor here. To provide a solution to this problem, we have introduced a methodology of a Bangla virtual assistant 'Adrisya Sahayak (Invisible Helper)'. As a desktop application, it is especially for Bangla speaking visually impaired persons to reduce the difficulties in utilizing a computer, peripheral devices of the computer and home appliances. Bangla is one of the most spoken languages globally and the native language of Bangladesh and a few states of India. A big portion of them is illiterate and not capable of communicating in English. That's why existing virtual assistants in the English language will be tough to use by them. The current user-independent system presented here will provide the opportunity to perform several tasks of a computer in the indoor environment via Bangla human voice commands.**

***Keywords*—**virtual assistant; Bangla speech recognition; visually impaired; human-computer interaction.

I. Introduction

The computer is vital in today's atmosphere. Our society is a global one, becoming more so with each passing day because of the rise of the internet and computers. Nowadays, we cannot think of our life without computers, and the fact is, their presence is essential that nothing can be replaced with them. Amelioration in technology transforms the way of communication with computers. One of the most convenient approaches to communicate with computers is the human voice command. It makes a comfortable medium to operate a computer for people with a lack of knowledge as well as to people with unwanted physical disabilities. Numerous researches have been done on speech recognition based virtual assistants in the English language since the past decade. The upswing of various virtual assistants such as Google's Google Assistant, Microsoft's Cortana, Apple's Siri, Amazon's Alexa, and Bixby by Samsung create an illusion with amazing functionalities for computer device users. However, there are very few virtual assistants in the Bangla language, but no virtual assistant is developed for the visually impaired who speak Bangla. There are 265 million natives and non-native speakers worldwide, and Bangla obtained 7th position among 100 most spoken languages worldwide [1]. Besides, the English Language's limited usages cause unconventional English virtual assistants' utilization among Bangla Speaking people. Developing an effective hands-free virtual assistant for visually impaired native Bangla speakers is our research's main purpose. It is facilitated by providing cozy computer use and accessing peripheral devices used in regular day-to-day life. In this application model, we use an Application Programming Interface (API) of Google, namely Google Web Speech API. It ensures that speech recognition and speech synthesis in the desktop application is performed smoothly. This API has a default API key that is hard-coded into the Speech Recognition library, enabling developers to access it without authentication [2] easily.

II. Existing Works

Several English speech recognition-based voice assistant systems exist for visually impaired persons, but very few of them are Bangla language supported. Prince Bose et al. [3] described a voice-controlled system in the English language for the blind. Email sending and receiving, news portal reading as well as weather forecasting functionalities were implemented. Google Speech API was used for providing speech recognition capability. "ABYS" [4] a voice-controlled desktop application for the English language developed using Microsoft's Speech Application Programming Interface (SAPI) for offline speech recognition along with Bing Speech API for long-phrase detection. The primary goal of that application was to reduce the hassle of visually impaired while using a computer and accessing home appliances. The system's drawback is Bing Speech API is not free, and home appliance accessibility is not wireless [5].

Ezekiel Marvin [6] developed a digital assistant focused on the visually impaired in recognizing English Text on real-world objects and providing audio feedback. Besides, this system performs voice interaction with the user, but the user interaction model is not well developed. In 2019, Dipankar Gupta et al. [7] presented a digital personal assistant using Bangla voice command and face detection for handicapped people. In that work, the cross-correlation technique, including the Bangla voice command, executes tasks defined by the pre-set command. The facial movement-based mouse cursor controlling system enabled a user to access the

computer more comfortably. A shortcoming of the system is it performs based on predefined command only. That's why it is unable to deliver results based on undefined voice command. The research conducted by Syed Mohidul Islam et al. (2019) [8] is very close to our proposed system. It introduced a development methodology of Bangla virtual assistant 'Adheetee (Erudite)' for smartphones and personal computers, which can perform various kinds of smart devices' functionalities. They used different algorithms to identify and respond accordingly for respective commands. Also, Google's Speech to Text API was used for speech to text conversion. Only inside the computer functions were implemented, but the home automation system or electronic device accessing capability is not included.

Another study [9] describes a personal assistant's development that assists users to interact with household appliances via speech and gesture. This English speech recognition-based system also enables a user to obtain some search results through web scraping. Developing 'Olivia,' a virtual assistant to make an ordinary home into a smart home, was the aim of the paper [10]. This system is enabled with the integration capabilities of Electronic home appliances in a limited scale. Besides, Limited functionalities like weather update, telling jokes, playing songs, or performing various calculations depending upon English voice command was implemented.

The authors of the paper [11] had proposed an Internet of Things(IoT) based smart home and a blind stick using Bangla voice command to control all the electronic components such as an electric lamp, TV, refrigerator, washing machine, etc.

III. Proposed Model

Our system is aimed at helping the visually impaired accessing online facilities at the initial stage. So, it is expected to ensure the availability of internet connectivity. If the internet connection is available, Google Web Speech API will be utilized. In the design mode, a microphone takes Bangla voice command as input and sends it to the API. Speech Recognition Engine of the API converts the recognized speech into Text.

After that, the Intent recognition system matches the converted Text with our customizable Bangla command database. If the Text matches with any command in the database, it operates program execution specified for it and provides speaker output after the speech synthesis. If the text equivalent of the voice is not matched with any of the existing commands, it passes the string towards a search engine via the default browser. It displays the output alongside the speaker generated output. The proposed model is shown in Figure 1.

A. Model Implementation

In this section, the operational output of our proposed system associated with different commands has been described. In our system, seven different sets of modules are facilitated for visually impaired people. The models are selected as per the user's voice command (Figure 2).

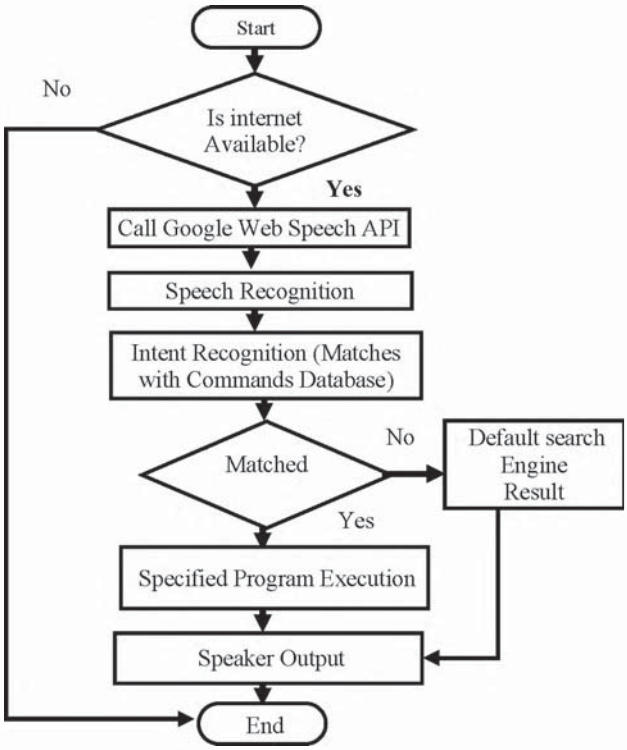

Figure 1: Proposed System's Execution Flow Chart.

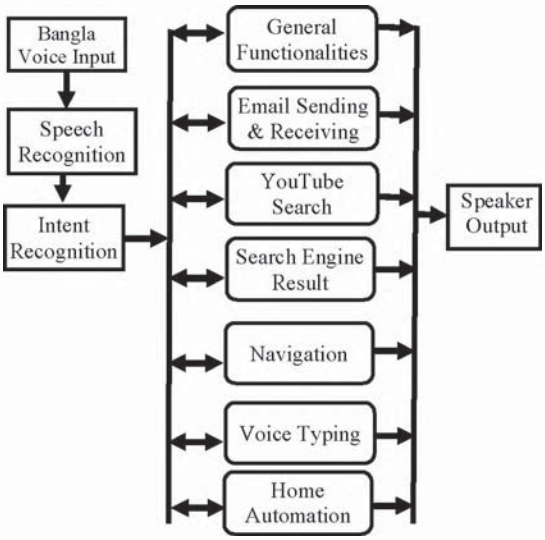

Figure 2: Application Module selection based on voice commands.

1. General Functionalities

Our system can perform general functionalities such as accessing the installed application on the Operating System (OS), setting the alarm, date and time information, getting weather updates, taking selfies, etc. There is a label on the top of the Graphical User Interface (GUI) that denotes the current status of the system, 'শুনছি বলুন' (Listening Say) while

speech recognition is continuing and 'ধারনা করছি' (Thinking) during the time of performing the desired operation. User responses and system responses are displayed in the GUI along with two labels such as called 'আপনি বলেছেন' (You said) and 'অদৃশ্য সহায়ক' (Adrsya Sahayak), respectively. (Figure 3). For example, in Figure 3, a Bangla command such as 'আজ বাংলা মাসের কত তারিখ' (What is the date of Bengali month today) was recognized and executed.

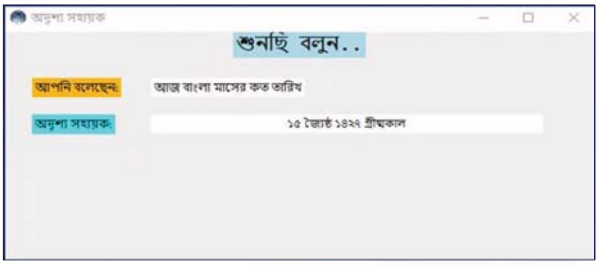

Figure 3: Execution output for a Bangla command such as today's date (Bengali months).

To access any installed application user needs to give the command name of the application and adding a phrase 'ওপেন করো' (Open it) or 'খুলো' (Open Up) in the end. Figure 4 shows a user request, sas 'মাইক্রোসফট ওয়ার্ড খুলো' (Open the Microsoft Word).

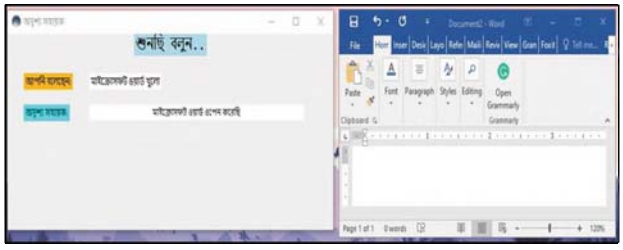

Figure 4: Application access (Microsoft Word) installed in the Operating System.

2. Email Sending and Receiving

Emailing is such a necessary work of our daily life that it is really hard to find a day without accessing mail once in a day. To send an email user needs to generate command 'ইমেইল পাঠাবো' (I will send an email) or 'ইমেইল পাঠাতে চাই' (I want to send an email). Then the system will start formatting email. (Figure 5). If the email is sent successfully, it will make speaker output as 'ইমেইল সফলভাবে পাঠানো হয়েছে' (Email Sent Successfully). Figure 6 shows the email sending result. The voice commands' ইমেইল পড়ে শোনাও' (Read and listen to emails) or 'ইমেইল ইনবক্সে যাও' (Go to the email inbox) takes the user to the most recently received email's body and generated the speaker output (Figure 7).

3. YouTube Search

If the user searches for any YouTube video, the requesting command would be the name of the video and 'ইউটিউবে চালাও' (Play on YouTube) or 'দেখাও' (Show on YouTube) at the end of the requested string.

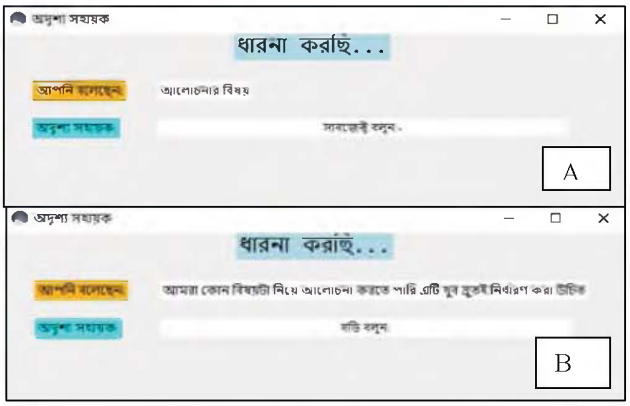

Figure 5: Email formatting (A. Subject and B. Body).

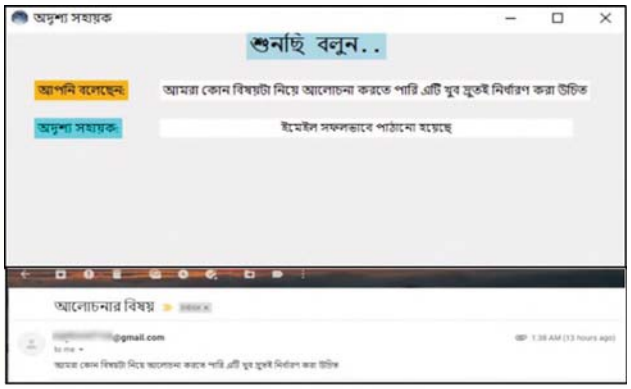

Figure 6: Email sending results.

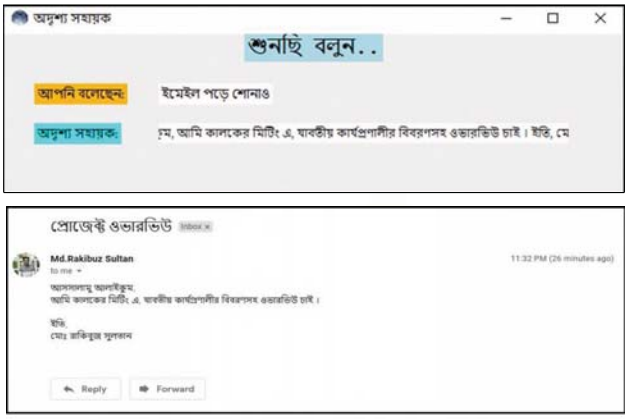

Figure 7: Email readout.

Figure 8 shows a Youtube Search result after 'স্বপ্ন যাবে বাড়ি আমার ইউটিউবে চালাও' (Sapna Jabe Bari Amar play on Youtube) command. Our system processed the command string at the Intent recognition stage, tracked the searching keyword, and played the topmost video on the search output list via default browser on the Operating System without putting the keyword on the YouTube search bar.

Figure 8: YouTube search result.

4. Search Engine Result module

The feature of searching any unknown term with the Bangla command is included in the proposed system. It initiates the result summary of seeking keyword on GUI and makes a speaker output from the search engine result. Google search result by this system is shown in the following Figure 9.

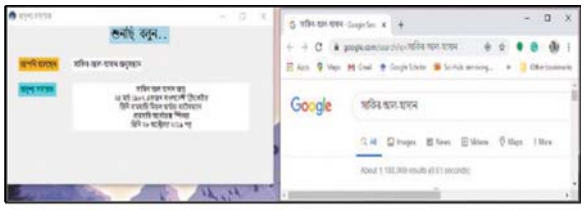

Figure 9: Google search result.

5. Navigation

Getting a general overview of distance and route direction between two separate places is commonly used in our daily lives. Our system enables its user to get map direction using Bangla voice command. For example, the command 'আদাবর থেকে যমুনা ফিউচার পার্ক যাওয়ার রাস্তা দেখাও' (Show me the way from "Adabor" to "Jamuna Future Park") command was processed in Figure 10 and deliver output on the default browser.

6. Voice Typing

This system sends the recognized Bangla speech to the desired cursor position. Users can utilize this module instead of typing Bangla in any application such as Facebook, Microsoft Word, Notepad, etc. Figure 11 shows the system is performing Bangla typing on Microsoft Word through voice. The Bangla command from a user was 'ভয়েস টাইপ ওপেন করো' (Voice Typing). In response to the command, the system enabled voice typing mode, and the reply was 'ভয়েস টাইপিং মোড চালু করেছি, কার্সর এর অবস্থান থেকে টাইপ শুরু হচ্ছে' (I have started voice typing mode, typing begins from the position of the cursor).

Figure 10: Desired direction output on the Google Map.

7. Home Automation

Being able to automate a series of regular home tasks can improve productivity greatly. Additionally, it makes the process handy as common household activities have centralized control. Our proposed Home Automation system architecture is represented in the following Figure 12.

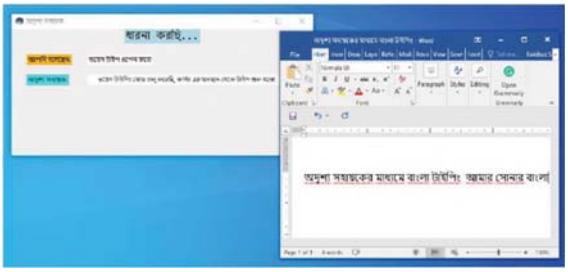

Figure 11: Voice Typing implementation.

Figure 12 encapsulates that all the electrical devices were operating in the concurrent environment. All of them are connected with the Microcontroller. The Wi-Fi module associated with the Microcontroller can initiate a private

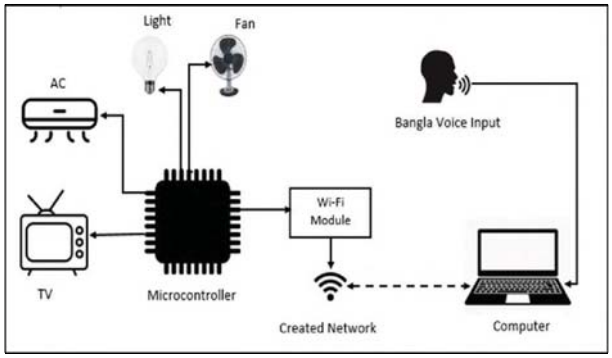

Figure 12: Home Automation System Architecture.

wireless network to accept the computer's requested command and forward the returned signal to that computer.

Figure 13 exemplifies the implementation of the proposed Home Automation Architecture and systems GUI and a lighted lamp.

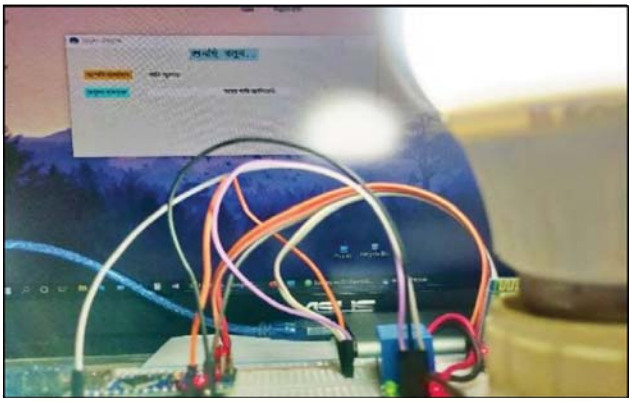

Figure 13: Home automation output.

IV. Performance Study

A performance study was conducted to determine the system's capability. We have tested our six different sets of operations concerning 1185 test cases. A total of 45 visually impaired people (Table 1) participated in the survey between January and September 2020. It was difficult to locate completely blind people willing to participate in the experiment. So, we focused more on the participants having blurred vision (Short distance vision problem).

Table 1: Statistical distribution of the participants

Gender count	Age range	Blurred vision	Blind
Male (36)	30-54 years	31	5
Female (9)	28-55 years	7	2

Table 2 shows the number of trial cases for different functional modules and the effectiveness of the commands given over microphone and successful execution. From the table 2, we can see that the average accuracy of our system is 93.3% and YouTube Search operation indicating a maximum of 96% for 175 trial cases. Noise in the environment, internet speed, and miss pronunciation create failure in a distinct set of operations.

A. Time requirement of the functionalies

Figure 14 shows an overview graph of the average response time for the six operation sets. Overall, most of the General Functionalities tasks were based on either predefined commands or the user preference's commands. This operation set takes a minimum amount of seconds among all the groups. Video quality and latency in internet connection were the reason for the highest and second-highest average time consumption to respond to the YouTube Search and Search Engine Result, as well as Navigation operation, sets, respectively.

Table 2: Average System Accuracy over various functionalities

Operation Name	Number of Trial Cases	Success	Failure
General Fun ctionalities	270	250/270≈ 92.6%	20/270 ≈7.4%
Email Sending & Receiving	30	28/30 ≈ 93.4%	2/30 ≈6.6%
YouTube Search	175	168/175 ≈96%	7/175 ≈4%
Search Engine Result	350	310/350≈ 88.5%	40/350 ≈11.5%
Navigation	150	144/150≈ 94%	6/150 ≈6%
Voice Typing	130	123/130≈ 94.6%	7/130≈ 5.4%
Home Automation	80	75/80≈93. 7%	5/80 ≈6.3%
Average System Accuracy		**93.3%**	**6.7%**

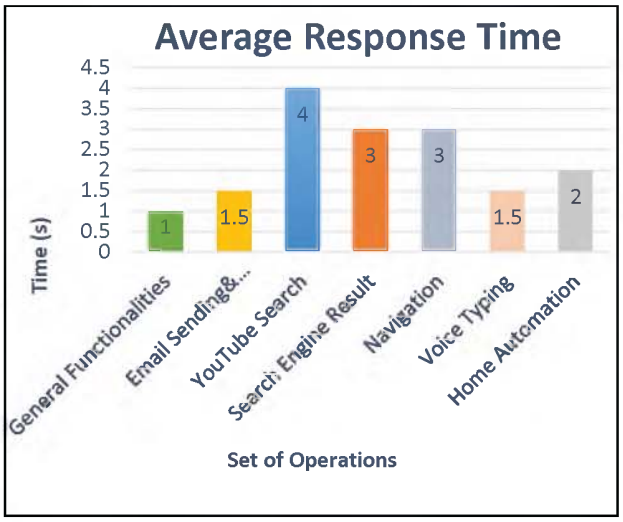

Figure 14: Average response time for operation sets.

The average response time of the entire system, which is 2.28s constructed from the six operation sets' average response time.

B. Comparison with similar systems

We have compared our system with systems available online in terms of average system accuracy. The work 'ABYS' proposed by Sultan et. al. [4] presented a similar system with limited functionalities in English language for the Bengali speaking people. Our system in Bangla voice and Bangla interface performs much better in terms of

accuracy. Figure 15 shows our position with [4] in terms of the accuracy of the functionalities.

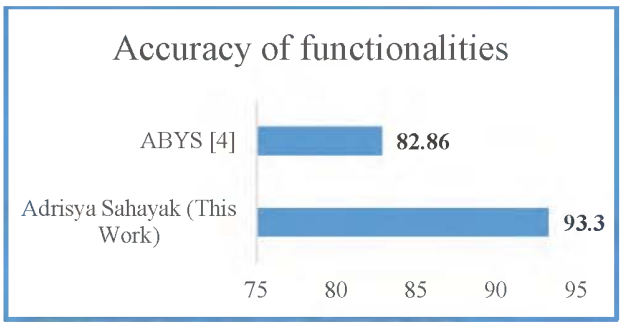

Figure 15: Comparison with similar Systems.

V. Conclusion and Future Scope

The motivation of our work came from the current state-of-the-art in assistive technology regarding Bangla speech recognition as well as the implementation of a Bangla virtual assistant for visually impaired people to use diverse functions of a computer and access integrated electrical devices associated with it. The proposed system is capable of interacting with the user via Bangla oral communication. Our implementation strategy and experimental analysis demonstrate a satisfactory outcome of this work. One of the key reasons behind the effective functional accuracy, in general, was the use of native language in the voice recognition component compared to using the English language.

'Adrisya Shayak' is not a perfect system. It is presented as an outcome of ongoing research work and updates, and the modification process of the work is still going on. We thought, presenting the current work will create interest among the research community to develop similar systems focusing on Bangla speech recognition. The presented system has some shortcomings. This system depends on the third-party API. Noise on the environment and internet speed may occur response delay while the system is operating.

In the future, developing a mobile version for user flexibility and developing API for minimizing dependencies are our concerns. The integration of the Machine Learning technique can generate productive replies much closer to human and decision-making capacity depending upon the surroundings.

References

[1] "Visual Capitalist", [Online] Available at: https://www.visualcapitalist.com/100-most-spokenlanguages.Accessed on May 24, 2020.

[2] "Speech Recognition", [Online] Available at: https://pypi.org/project/SpeechRecognition.Accessed on March 10, 2020.

[3] Prince Bose, Apurva Malpthak, Utkarsh Bansal, Ashish Harsola, "Digital Assistant for The Blind", 2nd International Conference for Convergence in Technology (I2CT), IEEE, April 2017.

[4] Md. Rakibuz Sultan, Md Moinul Hoque, "ABYS(Always By Your Side): A Virtual Assistant for Visually Impaired Persons" 22nd International Conference on Computer and Information Technology (ICCIT), IEEE, December 2019, pp: 1-6.

[5] "Cognitive Services Pricing", [Online] Available at: https://azure.microsoft.com/enus/pricing/details/cognitive-services/speech-api.Accessed on May 1, 2020

[6] Ezekiel Marvin, "Digital Assistant for the Visually Impaired", International Conference on Artificial Intelligence in Information and Communication (ICAIIC), IEEE, February 2020.

[7] Dipankar Gupta et. al, "A Digital Personal Assistant using Bangla Voice Command Recognition and Face Detection, International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON), IEEE Conference, Dhaka, December 2019.

[8] Syed Mohidul Islam et.al, "Adheetee: A Comprehensive Bangla Virtual Assistant", 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, May 2019.

[9] V. Chayapathy, G. S. Anitha, and B. Sharath, "IoT based home automation by using personal assistant," International Conference On Smart Technology for Smart Nation(SmartTechCon), IEEE, August 2017.

[10] Khattar S, Sachdeva A, Kumar R, Gupta R, "Smart home with virtual assistant using raspberry Pi", 9th International Conference on Cloud Computing, Data Science and Engineering(Confluence), IEEE, January 2019.

[11] M. W. Rahman, R. Islam, M. M. Hasan, S. Mia, and M. M. Rahman, "IoT Based Smart Assistant for Blind Person and Smart Home Using the Bengali Language," SN Comput. Sci., vol. 1, no. 5, pp. 1–13, 2020, doi: 10.1007/s42979-020-00317-6