Talkback Smart Mirror Using Magic Mirror and Google Actions

I. Khan[1], A. Jyotsna[2], H.V. Singh [3], S. Rasal[4*]

[1,2,3,4]Dept. of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College Of Engineering, Pune, India

*Corresponding Author: surasal@bvucoep.edu.in, Tel.: +91-8793000079

Abstract— Smart voice assistants have gained a lot of popularity over the years and have been integrated with numerous devices. This paper aims to focus on the contribution of the voice assistant to ease our lifestyles. Amazon Echo, Google Home are some of the already existing smart assistants. Currently, these systems cannot visualize the results. The main aim of this project is to integrate these assistants with ambient displays, which in turn would lead to increased efficiency, accuracy, and ease of access. The talkback mirror can be used for information retrieval, for conversations, and entertainment of the user. It will cater to the queries of users, without them having to lift a finger.

Keywords—Google Actions Console, Google Assistant SDK, Google Cloud Console, Magic Mirror, Mic, Natural Language Processing, Raspberry pi, Screen,Speaker, Two way Mirror

I. INTRODUCTION

In today's world many devices that we use in our daily lives utilize the voice assistants. These are present in our smart phones and in smart speakers that we can use in our homes. Many of the mobile apps and operating systems utilize them. Nowadays they are also being used in cars, retail, education, healthcare and telecommunication environments. For voice-enabled world the smart speakers have shown great potential. They make use of microphones and speakers to hear and communicate back with the user respectively. They are connected to the internet and advanced speech recognition software that understands and reacts to specific commands. The assistant actually deciphers the voice of the human and then acts to it by searching for the correct answer and reading it back to the user. The voice assistants don't actually understand what the user is saying they just wait and hear the 'wake up' word that the user is saying and then begin communicating with the server to complete the task. This is where Natural Language Processing is used i.e. for interpreting the human language.

The voice assistants provide us many hands-free tasks which is one of the main reasons why many people actually enjoy using them. These assistants can set appointments, make calls, set reminders, etc.

When Siri was introduced in iPhones in 2011, she changed the way we used our smart phones and technology. What Siri had changed for the smart phones, recently Alexa did for houses and gave rise to smart speakers. The markets for the smart speakers are blooming but the business world has been slow to adapt the technology. There is supposed to be growth in the number of people using the voice assistants. The importance of 'voice' is beyond doubt. Due to this the

companies are finding out better ways to make use of these voice assistants and integrating them into their customer services and markets.

The Talkback smart mirror shows user the news headlines, time and date, upcoming events and also listens to the user's request for the information he/she needs and produces the output of the request in the form of text as well as audio .This is made possible by using the magic mirror software to display the time, date and compliments on the screen. It uses the MMM-Assistant Mk2 module which is configured with the credentials created on the Google Actions Console and Google cloud console. It also uses the MMM-Hotword module which helps it understand the wakeup word that the user uses to give commands.

II. LITERATURE SURVEY

How people think about voice technology has really transformed over these 10 years. First voice assistants had limited use but now they are integrated in every part of people's lives. Listed down below are some popular voice assistants within the last 10 years:

2010- As a mobile app and then acquired by Apple- Siri
Siri was first introduced as a voice assistant mobile app for iPhones. Starting with the iPhone 4S, Apple integrated Siri as its voice assistant in every voice-capable product, including its current ecosystem of smart speakers and wearable technology. Siri became the first voice assistant, setting expectations and ideas about how voice assistants should operate for many people due to its early existence through iPhones. Siri has changed and expanded its abilities over the decade.

2011 – Google introduces Voice Search on the Homepage

In the 2000s Google started testing voice-based search technology on mobile devices, but in 2011 Google.com added a little microphone button for searching using voice. Applying the voice technology to the huge power of Google's search engine was an important step for the industry. Initially voice search was limited to the Google chrome browser, suddenly market had a huge chance to try out interacting with a website using their voice.

2013 – Microsoft's Cortana Voice Assistant

In 2013 Microsoft entered into the voice assistant arena and started off very strong. Cortana became part of Windows, Xbox, and other Microsoft creations, also took a spot in products built by other brands too. Microsoft ended Cortana's consumer-facing features and realigned it purely for business and enterprise efforts at the end of the decade.

2014 – Amazon Reveals Echo With Alexa Voice Assistant

Amazon built a smart speaker with its own voice assistant named for the Library of Alexandria. Alexa arguably is the most influential voice assistant now and is ahead of Siri or Cortana. Amazon is pushing the voice assistant into ever-growing array of smart speakers and other products, like wearables to cars.

2016 – Google Home And Google Assistant Debut

Google learned from its voice experiments and created Google Assistant, which challenged Alexa. Google Home and Google Assistant offer many features which are same, but with its integration into Google's larger technical ecosystem, such as with Android phones adds an additional lure. That became particularly true in the years after its debut when Google bought Nest and started creating a smart home network.
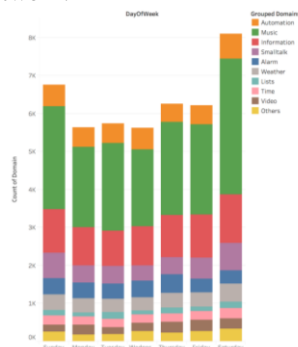

Figure 1: Google home use by day of the week, broken out by category. Use on weekends is significantly higher than use during week[1]

2018 – Samsung Releases Bixby 2.0 Voice Assistant

Samsung attempted to make space in the voice assistant market in 2017 when it debuted with Bixby. It took only six months for the company to decide to revamp it entirely. Bixby 2.0 didn't just solve some of the problems with the first iteration; it refocused on the voice assistant being more flexible for developers and more personalized for users than it was originally.

2019 – More Than 3 Billion Voice Assistants Are In Use

Billions of voice assistants are in use today. Smart speakers and smart phones are the most common ways to interact with the voice assistants. The ways people use voice assistants are growing quickly as the options for interacting with them. The stilted orders to make a phone call repeated until understood are now a casual conversation to order groceries, turn off the lights, and start the car.

The last year has seen debates over privacy with regard to voice assistants and issues of security and safety are not going to fade away. The next decade of voice assistants may include as much change as possible.

III. LIMITATIONS

With increasing popularity of smart voice assistants and increased user inclination towards it, there arise the concerns about privacy and security. It is difficult to ascertain the exact user of the device for the given command.

1. DATA COLLECTION AND PRIVACY

Smart speakers are designed in a way that these are always active and listening for the wake word.
Once the wake word is detected the speaker starts recording and the capture audio is sent to the cloud for processing and apt response is sent back to the user. Audio is not sent to cloud for processing until the wake word is detected, yet it gives an impression to users that the assistant is always listening and essentially recording everything.

Reports have emphasized on the tendency of the assistants to wrongly detect the wake words and start recording and processing of the audio. Voice assistants can easily be triggered by homophones, there have been numerous instances where Apple assistant Siri has been activated by "Hey, seriously" as it sounds similar to "Hey, Siri-ously", these unintentionally recorded snippets are referred to as false accepts.

While the LED starts flashing once the recording is initiated, at times users don't tend to notice the flashing, proving this notification technique ineffective several times.

The impact of false accepts on user privacy could be grave if the captured audio is processed on unauthorized servers and application. False accepts can lead to potential security threats.

2. SECURITY

While voice assistants have influenced the user's relationship and interaction with technology, but it has also triggered security concerns. The crux of security concerns includes access to the assistants by unauthorized persons or devices and executing the commands.

- Currently the speaker can be accessed by the supervised users, however these assistants are

vulnerable to attacks by unsupervised users having physical access to device. They can easily modify the device and the settings. Voice assistants implement speech recognition techniques to process the commands, differentiating between the commands in genuine human voice and pre-recorded commands is still a challenge for them.

- There have been numerous incidents where children have been able to make purchases with the help of smart speakers, without the knowledge of parents. The voice assistant although asks for confirmation of the order to prevent accidental shopping but if the child has the access, nothing can be done to restrain them.
Another way to trigger the speaker is through ultrasonic attack. The attacker needs an extra speaker, amplifier and he should be in the range of audibility. This kind of attack was carried out against such devices in 2015, electromagnetic waves were picked up by the cables and commands were triggered.

- Voice assistants can also take commands from other devices such as TV or radio. For example, in April 2017, Google home devices were triggered by advertisement by Burger King which mentioned "OK Google, what is whopper burger?" in a calm voice.
Another major case of smart speakers getting triggered by external influences was witnessed in September 2017, when devices were activated by an episode of South Park.

These are the major security setbacks faced by the smart assistants till date. In additional to these vulnerabilities, the assistant should be connected to secure network.

3. INTEGRATION OF EXCESSIVE COMMANDS

Voice recognition implements natural language processing to process the commands. The model works well as long as it is expected to perform specific tasks. Interpretation of ambiguous commands is still a challenge and could lead to chaotic results.

The assistant is generally used to perform 3-4 specific tasks which includes automation, information retrieval, setting remainders and alarms and playing music. Problem with integration of additional commands is the user would not be able to remember each and every command. Additional command could increase complexities and create confusion for the user.
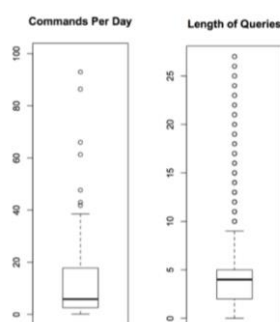

Figure 2 :Commands issued per day to Google home (left) and average length of commands in words(right) [1]
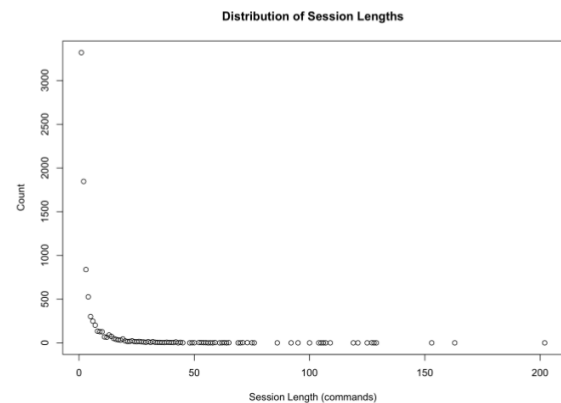

Figure 3: A distribution of commands per session across all interactions with device. 39% of all sessions contained only a single command[1]
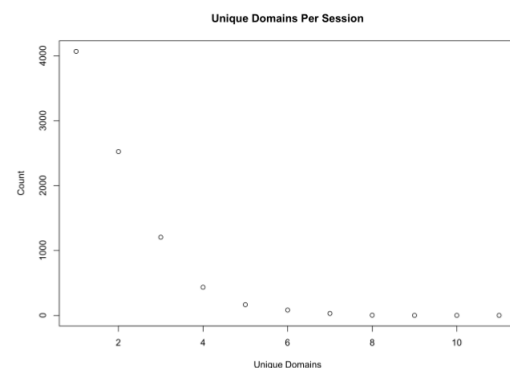

Figure 4: Number of unique domains used within a session. 48% of all sessions only involved a single domain[1]

4. CONVERSATIONAL CONSTRAINTS

Voice assistants mostly support one-way conversations, they don't talk back asking for the context of the command or for any extra instructions.

These assistants are unable to initiate conversations, answer random questions or keep track of the conversations, as a result these devices lag conversational skills. These assistants are unable to maintain the state of the commands and interpret their context.

IV.DESIGN ADVANCEMENT

1. DOMAIN DISCOVERY

According to research and survey regarding the usage trend of smart speakers, the fact that it generally used to select and play music, set alarms and remainders, answer the questions, is very evident. Users neither tend to explore other functionalities nor do they tend to explore the updated commands. The smart speaker could help users to discover additional skills as per the interactions they make and suggest about the related features. For example, the device could suggest the alternate updated commands to users for recurrent actions like setting up alarms, remainders and timers.

2.   CUSTOMIZED COMMANDS

User's interaction with the voice assistant is very one way. Repetitive commands are used for frequent actions. These commands are generally long sentences or frequent phrases. These recurrent commands can easily be customized and shortened which would make interaction easier and personalized.

Basic commands can be replaced by customized and improvised commands by engaging in dialog to rename the recurrent commands. This will help create user's shortcuts.

Command	Percent
stop	7.3%
what time is it	2.7%
pause	1.1%
how much time is left	1.1%
pause TV	0.8%
play	0.8%
skip this song	0.8%
volume up	0.8%
tell me a joke	0.8%
what's the temperature	0.8%
resume TV	0.7%
volume down	0.7%
next song	0.6%
next	0.5%
turn on kitchen	0.5%
what's the temperature outside	0.5%
stop TV	0.5%
turn on table	0.5%
turn off kitchen	0.5%
turn off living room	0.4%

Figure 5: Top 20 commands made to Google Home[1]

3.   INTEGRATION WITH AMBIENT DISPLAYS

The capability of smart voice assistants has improved after its integration with the ambient displays. This timed multi-modal interaction is capable of anticipating user's requests and requirements and providing the relevant information, which is visible on the display.

The ambient displays can be further updated in terms of user interface, the way the voice assistant interacts with the display. Frequently used features and additional functionalities related to these features should be listed on the screen, brief information can also be displayed on the screen to help user discover additional skills and domains.

V. METHODOLOGY

Our system makes use of Google Actions Console and Magic Mirror software. The hardware includes raspberry pi 3b+, 5 inch screen, mic, two way mirror and speaker. The drawback of the smart speakers is that they only return the output of the requested query in the form of voice. Hence to overcome this drawback we have integrated a screen in our project. When the user will give a command to the assistant it will process it and display its output on the screen with the help of the Magic Mirror software.

Below mentioned are the softwares used in the project:

- **Magic Mirror Software:**

It is a open source smart mirror software. It provides various installable modules that can be used for building a personal assistant.

- **MMM-Assistant Mk2:**

It is a module with the help of which you can do the following using your voice:

i.   Control your other Magic Mirror modules

ii.   Provide Google Assistant Methodology

- **MMM-Hotword:**

It detects the hotword using snowboy. We use this module to wake voice assistant or can also be used to give a command to other module.

- **Google Actions Console and Google cloud console:**

Google Actions is basically a web-based tool that is used to develop actions. It is used to manage Action registration, deployment, analytics and configuration. Google cloud console helps us deploy, scale, and diagnose production issues in a simple web-based interface

We used the following hardware:

- **Raspberry pi 3b+:**

Raspberry pi is a series of small single-board computers which runs on Linux operating system. It is a microcontroller that can be easily connected to computer, monitor etc. We have used Raspberry pi 3b+ that has a faster 64bit 1.4 GHz quad core processor , 1GB RAM, faster dual-band 802.11 b/g/n/ac wireless LAN, Bluetooth 4.2, and significantly faster 300Mbit/s Ethernet.

- **Mic:**

This is used as an input device for the Talkback Smart Mirror. It is used to capture the hotword that the user calls out to wake up the assistant.

- **Speaker:**

This is used as an output device. It is used for voice output for the user's request.

- **HDMI Screen:**

HDMI Screen is used for displaying the output that the user has asked the assistant.

VI.   WORKING

The input is given by the user in the form of voice commands with the help of the mic. The MMM-Assistant Mk2 is installed on the raspberry pi with credentials for accessing the Google assistant services. When the user sends the voice commands through the mic and asks for a service. The device then sends the model and device instance identifiers so that the service can determine how to respond in the best way to the context.

The service uses Automatic Speech Recognition (ASR) to find out what words were actually spoken by the user. It also uses Natural Language Processing to analyze and find out the meaning of the request from these words. It then searches for the information asked by the user and

responds with the text that is spoken back to the user along with the information asked for by the user.

This text and the received output is displayed on the screen with the help of the magic mirror software on the screen that is projected on the two way mirror. And the text that is spoken back to the user is heard with the help of the speaker attached to the raspberry pi.
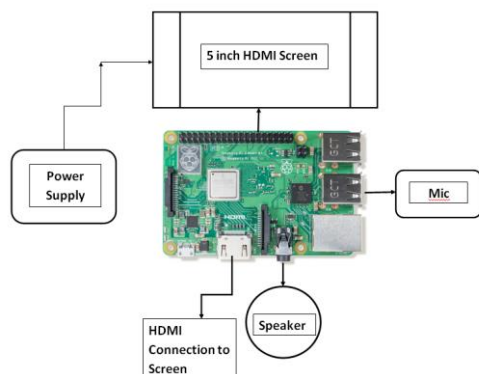

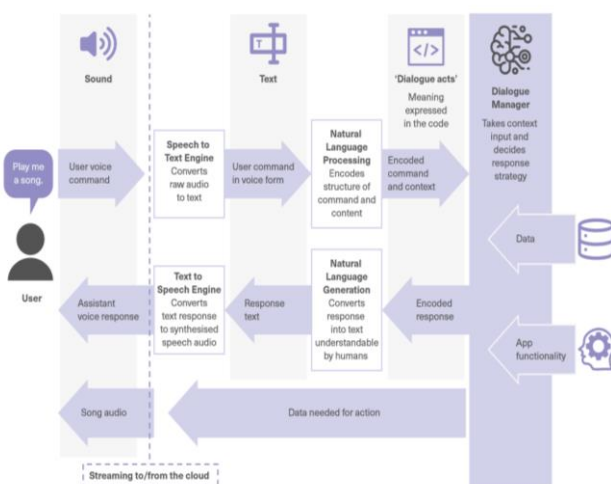Figure 6: Working diagram of Talkback Mirror

Figure 7 : Adapted from diagram by Prof Verena Rieser, Heriot Watt University [4]

VII. FUTURE OUTLOOK

The capabilities of voice assistants are bound to improve with advancement in technology. Embedding of assistants with various appliances would lead to wider and enhanced skill sets and more natural interactions.

Integration of third-party applications will enhance accuracy and precision of the assistant. User would be able to avail features like customized and improvised commands. Enhancement can be expected in terms of conversational features with the help of context processing and agent integration. Assistant can provide health advises if integrated with medical domain. The assistant would be able to perform sequence of actions by detecting the state of the devices.

Edge computing with voice printing and privacy first architecture will improve the privacy and security of the speaker. Privacy and security could be further enhanced by providing the user the facility to review what is being shared. This will provide user with greater control over their data.

User can experience more personalized interaction with help of the interaction history, which would require the assistant to maintain the history of the interactions for every user.

More customized notification, suggestion, recommendations and replies can be expected with the help of sentiment analysis of interaction history.

VIII. CONCLUSION

The major problem associated with existing voice assistant systems are Siri and Cortana can only be used on phones, tablets, or computers. Amazon's Echo, Google Home both serve as speakers which are using the talkback feature while interacting with the users. The commands used for interactivity are repetitive. Without the presence of any display, it might be difficult for users to determine the accuracy of results and to comprehend the desired information.

The talkback smart mirror will help users to have the access to information on the mirror. Users will have the opportunity to determine the validity and credibility of the obtained results. The ambient display will contain the date, time, and other essential information such as temperature details, this will help users to explore other functionalities rather than using the system for the same functions. This talkback mirror uses Raspberry Pi 3b+ to provide better connectivity, HDMI screen for display, Magic Mirror Software, and associated modules for building personal assistant.

REFERENCES

[1] Bentley, F., Luvogt, C., Silverman, M., Wirasinghe, R., White, B. and Lottridge, D., 2018. Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), pp.1-24.

[2] Wueest, C., 2017. Everything You Need to Know About the Security of Voice-Activated Smart Speakers. Symantec.(Nov. 2017).

[3] Norberg, F. and Yildirim, Z., 2018. The implementation of Voice Command in Smart Homes.

[4] Rieser, V. and Lemon, O., 2011. *Reinforcement learning for adaptive dialogue systems: a data-driven methodology for dialogue management and natural language generation*. Springer Science & Business Media.

[5] Newman, N., 2018. The future of voice and the implications for news.

Authors Profile

Ms. Iffat Khan is currently pursuing Bachelor of Technology from Bharati Vidyapeeth (DU) College of Engineering, Pune.

Ms. Amrita Jyotsna is currently pursuing Bachelor of Technology from Bharati Vidyapeeth (DU) College of Engineering, Pune.

Mr. Harsh Vardhan Singh is currently pursuing Bachelor of Technology from Bharati Vidyapeeth (DU) College of Engineering, Pune.

Mr. Suraj Rasal is currently working as an Assistant Professor in the Department of Computer Engineering, Bharati Vidyapeeth (DU) College of Engineering, Pune. He has published more than 30 research papers in various international journals.