ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.10

Unified Voice Assistant and IoT Interface

Pradeep Doss[1], Ankit Pal[2], Kripa Jaba Sing Paul[3]
Assistant Professor[1], Student[2, 3]
Department of ECE
SRM IST, Ramapuram, Chennai, India

Abstract:
Voice assistants are becoming progressively smarter and more astute because of improvements in computerized reasoning innovation. While their principle work is to react to directions, in doing as such, they additionally learn. The more a man interacts with voice-actuated gadgets, the more patterns and examples the framework recognizes depending on the data it gets. At that point, this information can be used to decide client inclinations and tastes, or, in other words term offering point for making a home more quick witted.But the shortcomings and exclusiveness in services offered by the service providers in this sector has ledplethora of options and gadgets choking up user space and money. And it has increasingly become difficult to adopt and manage new age technologies. Hence there is a need to develop a universal system, so the proposed solution is to develop a future expandable platform which can run multiple voice assistant services at the same time with a cost-efficient device that allows easy and seamless upgradation to smart homes and buildings.

Keywords: IFTTT, ESP8266 Module, Linux OS, Python

1. INTRODUCTION

A. Motivation

1. Field

The present technologies of voice-based interactive controls relate generally to web browsers and search engines and, more specifically, to user interfaces for web browsers using speech in different languages. The unified voice assistant and IOT interface is a human-machine interface which can provide an easy and seamless integration of services and IOT technology into a single voice based interactive device.

2. Description

Currently, the Internet provides more information for users than any other source in the world. However, it has been often difficult to find the information one is looking for with ease. In response, search engines have been developed to help locate desired information. To use a search engine, a user typically types in a search term using a keyboard or selects a search category using a mouse. The search engine then searches the Internet or an intranet based on the search term to find relevant information. This user interface constraint significantly limits the population of possible users who would use a web browser to locate information on the Internet or an intranet, because users who have difficulty typing in the search term in the English language (for example, people who only speak Chinese or Japanese) are not likely to use such search engines.

When a search engine or web portal supports the display of results in multiple languages, the search engine or portal typically displays web pages previously prepared in aparticularlanguage only after the user selects, the desired language for output purposes using a mice or keyboard.

Recently, some Internet portals have implemented voice input services whereby a user can ask for information about certain topics such as weather, sports, stock scores, etc., using a speech recognition application and a microphone coupled to the user's computer system. In these cases, the voice data is translated into a predetermined command the portal recognizes in order to select which web page is to be displayed. However, the English language is typically the only language supported and the speech is not conversational. No known search engines directly support voice search queries.

Since the present technologies of voice-based interactive controls relate generally to browsers and search engines and more specifically to user interfaces for web browsers using speech in different languages. The unified voice assistant and IOT interface is a human-machine interface which can provide an easy and seamless integration of services and IOT technology into a single voice based interactive device. This has the potential to eliminate the use of multiple products and expensive services for the same and enables better and efficient user experience and affordability.

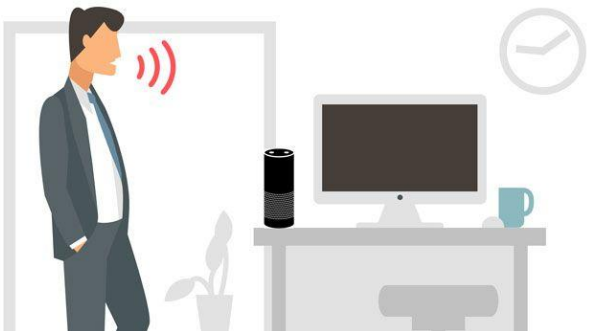

Fig. 1.1. Voice based System

2. SYSTEM DESIGN

The goal of the unified voice assistant and IOT interface project was to have a microcontroller board to be as small as possible, yet not compromising on the performance of the system. The system should integrate seamlessly and provide faster and power efficient performance.

This is achieved by the use of powerful yet power efficient on-board computer system and reducing the load on the local system as much as possible. It is preferred throughout the design to offload the heavy data processing requirements to off-site systems.

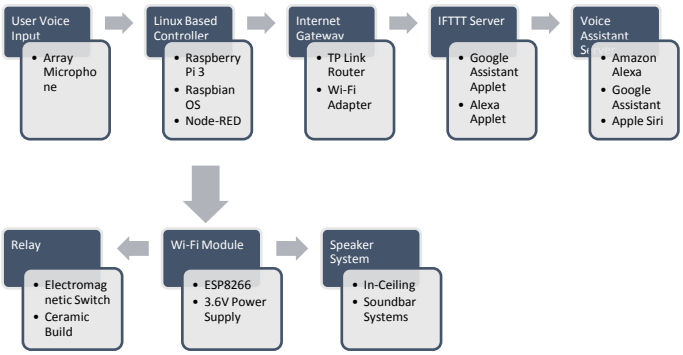

Fig. 2.1 System Model Block Representation

A. On Board Computer System

The Raspberry Pi hardware has advanced through a few forms of upgrades that varies in memory limit to fringe gadget bolster to support better integration.

Raspberry Pi Model B and B+, Model An, A+, and the Pi Zero are comparative, yet come up short on the Ethernet and USB center point segments. The Ethernet connector is instead associated with an extra USB port. In Model An, A+, and the Pi Zero, the USB port is associated specifically to the framework on a chip or system on a chip (SoC). On the Pi 1 Model B+ and later models the USB/Ethernet chip contains a five-port USB center point, of which four ports are accessible, while the Pi 1 Model B just gives two. On the Pi Zero, the USB port is likewise associated straightforwardly to the SoC, however it utilizes a smaller scale USB (OTG) port.

Fig. 2.2Raspberry Pi Model

The Raspberry Pi 3, with a powerful quad-core ARM Cortex-A53 processor, is portrayed as having ten times the execution speed than that of a Raspberry Pi 1 i.e. (3000 GHz). This was proposed to be very needy for the assignment threading and guidance set utilization. Benchmarks demonstrated the Raspberry Pi 3 to be roughly 80% quicker than the Raspberry Pi 2 in parallelized errands. The board runs Raspbian OS and Node RED and bash scripting is used to program the functionality of the system.

B. Internet Gateway

The unified voice assistant and IOT interface system is an highly network dependent system. A very stable and fast internet connectivity and good Wi-Fi signal strength is required throughout the area in which the system is being deployed. A highly efficient Wi-Fi router which works on IEEE 802.11 b/g/n/ac standard is preferred to achieve the best performance and efficiency. Adequate repeaters and signal amplifiers must be installed to gain the best range and efficiency. The Wi-Fi network with better network speed capability will provide the best results. If the on-board computer system does not support Wi-Fi an adapter can be used to extend its functionality.

C. IFTTT

If This Then That, also known as IFTTT, is a free web-based service that permits us to createsimple chains of conditional statements, called applets.An applet is activated by changes that happen inside other web services, for example, Gmail, Facebook, Telegram, Instagram, or Pinterest.For instance, an applet may send an email if the client tweets utilizing a hashtag or duplicate a photograph on Facebook to a client's document in the event that somebody labels a client in a photograph.

Services are the basic building blocks of IFTTT. They mainly portray a progression of information from a specific web server, for example, YouTube, amazon or even eBay. Services can likewise portray activities controlled with certain APIs, similar to SMS. Once in a while, they can speak to data as far as climate or stocks. Each service has a specific arrangement of triggers and actions.

Triggers are the "this" portion of an applet. They are the things that trigger the activity. For instance, from an RSS channel, you can get a warning or message dependent on a catchphrase or expression.
Actions are the "that" some portion of an applet. They are the actions that are the result of theof the triggers.

D. Sub Systems

- Microphone and Speaker System

The Microphone and Speaker System is the public addressing system, which is an electronic sound management system comprising microphones, amplifiers, loudspeakers, and related equipment. It increases the apparent volume (loudness) of a human voice or other acoustic sound source and provides high definition playback and interactive experience.

The DSP Group HDClear 3-Microphone Development Kit for Amazon AVS enables very low-power wake-on-voice operation and far-field voice recognition performance by using only three microphones. It is well suited for a broad range of applications, including the development of battery-powered devices such as portable smart speakers, IoT gadgets, and smart home devices as we require.
The solution includes the audio-centric DBMD5 processor that incorporates the DSP, HDClear voice enhancement signal processing algorithms, and Voice Activity Detector with three power operating modes (low power 1-microphone voice trigger, 3-microphone noisereduction, and 3-mic barge-in). This optimized architecture makes it possible for battery-powered devices to continuously monitor voice activity and detect the “Alexa” wake word or any other configured wake word while minimizing system-level power consumption. Features include:

- A reference client for the Raspberry Pi 3 is built using the AVS Device SDK
- HDClear algorithms are used for beamforming, noise reduction, and Acoustic Echo Cancellation

- The HDClear DSP Board comes with a dual-core DBMD5 processor as shown in fig.2.1

Fig. 2.3HD DSP board with dual -core DBMD5

- IOT Hardware Interface System

For Connectivity, we use ESP8266 module. This module allows microcontrollers to connect to a Wi-Fi network andallows it to make simple TCP/IP connections using Hayes-style commands. A relay arrangement acts as a digitally controlled switch.

The processor in the ESP8266 is the 32bit L106 RISC micro processor and runs at 80 MHz It has a 32 KB instruction RAM, 32 KB instruction cache RAM, 80 KB user-data RAM 16 KB system-data RAM.

The system works on the IEEE 802.11 b/g/n Wi-Fi network standards. The WEP or WPA/WPA2 authentication is used to restrict access and secure the data alternatively it also supports open networks

It has 16 GPIO pins which allows better expandability. The system runs on I²C, I²S interfaces with DMA, UART on dedicated pins.

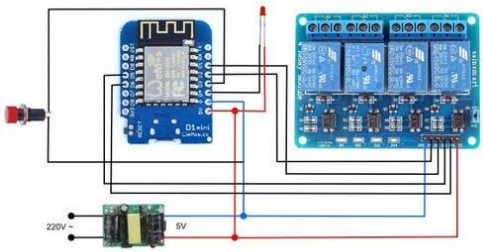

Fig. 2.4 ESP8266 and Relay Arrangement

3. MATHEMATICAL GOVERNING PRINCIPLES

A. Sound Intensity

$$I = P / (4 \bullet \Pi \bullet R^2)$$

(I=INTENSITY, P=POWER OF SOURCE, R=RADIUS)

The intensity of sound plays an crucial role in determining the effectiveness of the system. The users sound should e clear and loud enough for the microphone to pick that up and use it as an input. The microphone placement should be designed to identify the wake words and listen to the voice commands from the user.

B. Hidden Markov Modelling

Hidden Markov Model (HMM) is a measurable model in which the framework being demonstrated is thought to be a Markov procedure with unobserved states states. This model reduces the error in the audio signals by identifying the missing words using the already known data.

4. CONSTRAINTS AND DESIGN VARIABLES

A. Voice Interference

When two sound sources of very similar yet different frequencies meet at the microphone the phenomenon of beats is observed. Better acoustic properties of a room directly contribute to a better performing voice assistant system. Sounds like pops can be filtered out but might lead to involuntary actions. Proper microphone array placements influence the Quality factor.

B. Latency

Depending on the network speed the reaction time of the voice assistants can differ. The system can afford a maximum latency of 15 seconds. Lower latency provides better efficiency and better user experience. Better networks with faster internet access and good Wi-Fi signal strength all around the area under the system with very low signal loss can contribute effectively to decrease the overall latency and provide much better user experience.

5. EXPERIMENTALSETUP ANDTESTING

A. Setting up a Self-Test Room

Our experience and knowledge acquired has helped us focus on some standard attributes that make an individual test room compelling for voice execution assessment. The room necessities are like those characterized by the European Telecommunications Standards Institute (ETSI), with some customization. In particular, we are utilizing ETSI EG 202 396-1 V1.2.2, Section 6.1. The best part is that our proposal doesn't expect the tester to assemble a particular sort of room – We can empower self-testing in a run of the mill office room. These are the base necessities:

B. Room Size

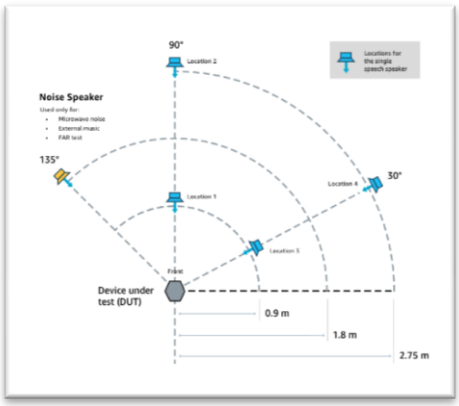

Fig. 5.1. Coverage range

We suggest a room something like 2.5 m x 3.5 m x 2.2 m. These measurements are in reference to usable space only. The coverage area and the equipment required for setting up the system is shown in the fig.3.

C. Room Treatment

Search for a space with end to end carpet cover on the floor and some acoustical damping (roof tiles) in the roof. In the event the room has a ton of windows or a substantial

whiteboard, consider covering these with an arrangement of shades to keep away from solid reflections by hard surfaces. We focus on a resonation time that is under 0.7 s and higher than 0.2 s. This is focused for the recurrence extend caught by the device, between 100 Hz and 8 kHz. For the technical specifics we concede to ISO 3382-3:2012. In the event that you are building a test chamber, it's better to consider 10 Hz neoprene isolators and bass snare boards in the corners.

*D.*The Noise Floor

To lessen the impact of undesirable clamor on the outcomes, the lowest acceptable commotion or noise of the test room ought to be under 35 dBA.

*E.*Equipment

An ideal setup requires 1 commotion or noise speaker, 4 normal speech speakers, your gadget under test and no less than a 0.5 m leeway between the dividers, test speakers, and your gadget. To control the yield of the speech speakers and commotion speaker you will need to get a multi-channel sound card, for example, the Roland Octa-Capture or RME Fireface.

6. CONCLUSION

Voice User Interfaces are ending up universally accessible, giving extraordinary chances to propel our comprehension of voice cooperation in a thriving cluster of practices and settings.

Numerous gadgets we utilize each day use voice assistants. They're on our cell phones and inside speakers in our homes. Numerous versatile applications and working frameworks utilize them. Also, certain innovation in autos, and additionally in retail, training, human services, and broadcast communications can be worked by voices.

From an availability viewpoint, voice controlled locally situated Intelligent Personal Assistants (IPAs) can possibly enormously extend our discourse collaboration past the old screen reader actions.

This system has the potential to greatly improve seamless interoperability of services and easy upgradation to smart buildings.

7. FUTURE PROSPECTS

Since the beginning of computing, UIs have been continuously evolving and becoming much more natural to interact with. The screen and console were one stage toward this path. The mouse and graphical UI were another. Touch screens are the latest advancement. The subsequent stage will no doubt comprise of a blend of augmented reality, signals and voice directions. All things considered, it is usually less demanding to make an inquiry or have a discussion than it is to type something or enter various points of interest in an online form. This applies to most users.

As voice assistants progressively wind up ordinary, this will likewise significantly affect marketing. As the case of web searches show, the assistant normally endeavors to give the best response to an inquiry and not ten distinct responses to browse from. Google's query items pages have officially created toward this path lately.Voice assistants will very quickly increment in significance in this regard. What's more, the fight for the best spots will be much harder than it is today. All things considered, focus will move from being displayed on the principal page of Google's query items to being the best of the list to be viewed as the best answer.

It remains to be perceived how publicizing can be incorporated into visual and voice assistants. Initial endeavors to do as such have had a tendency to befuddle clients. The most sensible and adequate form for this remains to be resolved.

The user experience can be further enhanced by adding visual screens which can make the system more interactive. Onething, without a doubt can be a breakthrough is if the assistant can read your feelings and act accordingly - the tech organizations that are building these capabilitiescanuse our voice to aide and become acquainted with us, examine our feelings and react with comparing human-like feelings. To what extent this will have go to get to the phase that we have an inclination that we're conversing with a human as opposed to a robot relies upon who you ask - five years, fifteen years, fifty - however with the speed at which technology progresses we can expect that sooner rather than later, Virtual assistants will have the capacity to perceive that you're in a specificemotional condition and react as needs be. Many virtual assistant services as of now has a Brief Mode that you can enact to get less talk, so we can perceive the amount to whichthis would be helpful if the assistant could check your mind-set and react accordingly.

REFERENCES

[1] Lee, Minsub, et al. "Method and apparatus for adjusting detection threshold for activating voice assistant function." U.S. Patent No. 9, 240, 182. 19 Jan. 2016.

[2] Byford, Roger Graham, et al. "Voice assistant system." U.S. Patent No. 8,255,225. 28 Aug. 2012.

[3] McConnel, Michel. "Alexa, How does Siri work? Voice Control Explained." Make Use Of, www.makeuseof.com.22 March 2016.

[4] "Waves, Sound and Light: Sound and Music." The Physics Classroom, www.physicsclassroom.com/calcpad/sound 14 October 2018.

[5] Aron, Jacob. "How innovative is Apple's new voice assistant, Siri?." (2011): 24.

[6] Berol, Dave. "How to create a self-test room and evaluate Voice Service Integration." Amazon Developers, www.developers.amazon.com.17 July 2018.

[7] Upton, Eben (14 March 2018). *Raspberry Pi Blog*. Raspberry Pi Foundation. Retrieved 2018-05-04.

[8] Emerick, Charles Thomas, et al. "Voice assistant system." U.S. Patent No. 9,171,543. 27 Oct. 2015.

[9] Binder, Justin, et al. "Voice trigger for a digital assistant." U.S. Patent Application No. 14/175,864.

[10] Ur, Blase, et al. "Trigger-action programming in the wild: An analysis of 200,000 ifttt recipes." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016.

[11] Rajput, Harshada, et al. "Implementation of Voice Based Home Automation System Using Raspberry Pi." (2018).

[12] Sonali Sen, Shamik Chakrabarty, Raghav Toshniwal, Ankita Bhaumik," Design of an Intelligent Voice Controlled Home Automation System," Department of Computer Science St. Xavier's College, Kolkata international Journal of Computer Applications (0975 – 8887) Volume 121 – No.15, July 2015.

[13] Tibler, Jan. "Alexa, What is the future of Voice assistants."DMexco, www.dmexco.com/stories/alexa-what-is-the-future-of-voice-assistants. 19 July 2018.

[14] Leggetter, Christopher J., and Philip C. Woodland. "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models." *Computer speech & language* 9.2 (1995): 171-185.

[15] Vorapojpisut, Supachai. "A Lightweight Framework of Home Automation Systems Based on the IFTTT Model." *JSW* 10.12 (2015): 1343-1350.

[16] Mehta, V., et al. "Smart and Interactive Home Using Raspberry Pi." (2018).

[17] Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77.2 (1989): 257-286.

[18] Varga, A_P, and R. K. Moore. "Hidden Markov model decomposition of speech and noise." *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*. IEEE, 1990.

[19] T. Anitha1, T. Uppalaiah2, "Android Based Home Automation using Raspberry Pi"1Assistant Professor, 2PG Scholar, Dept. of IT, GokarajuRangaraju Institute of Engineering and Technology, Bachupally, TS, India. International Journal of Innovative Technologies Vol.04, Issue.01, January-2016.

[20] Mi, Xianghang, et al. "An empirical characterization of IFTTT: ecosystem, usage, and performance." *Proceedings of the 2017 Internet Measurement Conference*. ACM, 2017.

[21] Peng, Chen-Yen, and Rung-Chin Chen. "Voice recognition by Google Home and Raspberry Pi for smart socket control." *Advanced Computational Intelligence (ICACI), 2018 Tenth International Conference on*. IEEE, 2018.