International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 594-599 www.ijaem.net ISSN: 2395-5252

# Iot Voice Based Command Execution and Speech Reconition System

Sulaiman Akilu Hamza
*School Of Computer Science,*
*Skyline University, Nigeia*
*(MSc Software Engineering (2676))*
*Dr. A. Senthil Kumar Dean, School of Science and Information Technology, Skyline University, Nigeria*

Date of Submission: 05-07-2025 Date of Acceptance: 15-07-2025

**ABSTRACT**
This study proposes a voice recognition-based control system that is both economical and efficient for elderly and disabled people. Without depending on a centralized Internet of Things (IoT) management system, the goal is to create an affordable speech recognition system that makes it simple to access devices placed in smart homes and hospitals. A multipurpose voice-activated smart home system using the ESP-32 as the wireless alternative is presented in the study. Voice commands are recognized by a specialized hardware component, and the database receives the natural language input. It analyzes information from the database on the recognition unit, decodes user-spoken commands, and manages household electronics.
Keywords – speech recognition, voice commands, iot, synthesizer.

## I. INTRODUCTION

Speech is a potent instrument that people can use to communicate needs and viewpoints. As a result, it makes sense that voice command machines have become increasingly popular for communication. The emergence of IoT applications[1] has led to the introduction of novel methods for sharing, interacting, and communicating from both a human and a machine standpoint. It may be less expensive to use voice commands for access. The Internet of Things (IoT) is expanding quickly and can be used to connect, manage, and survive items that are online [13].

However, some of the biggest obstacles to deploying IoT technology for daily use cases include device compatibility, security, and IoT device administration [2]. As IT has developed, we now live in a hyperconnected world where everything is connected and communicates via mobile devices and the internet. A key component of this network, which connects the real and virtual worlds, is machine-to-machine, or IoT, connectivity. Things in the real world are not necessarily distinct from the virtual world and can be controlled.

## II. LITERATURE REVIEW

The writing study has done a great deal of work on the hitherto under-considered role of voice in IoT and sees use cases and possible outcomes where voice thought not only improves the particular course of action but also makes financial sense.[1] [10]. Data and voice integration, such as VoLTE, into IoT applications has been the focus of The Trap of Things. This approach offers a flexible way to facilitate individual collaboration and communication and can provide a more intelligent user interface than standard methods, such as contact screens or data input with valid execution and taken into consideration as crucial human factors and costs.

This article uses Firebase as the cloud and the Android framework as part of a smart home computerization model. The Arduino Center point Mcu Wi-Fi Module's consolidation placed restrictions on all of the devices [2]. In order for the most recent computerization structures to communicate with their home machines, Paul Jasmin Rani et al. [3] suggested that they follow a specific set of guidelines or tactics. Clients leave the application as a result of these lengthy cycles. To address each ongoing problem and change the game strategy, the suggested solution employs voice commands to communicate with home computers via a wireless network [11].

It was Normal Language that handled these speech commands. Taking care of clients encourages them to use the application more by

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 594-599 www.ijaem.net ISSN: 2395-5252

giving them a more grounded interface. Additionally, it eliminates the mundane task of actually operating household computers. In order to control devices in magnificent homes or smart facilities, this article suggests a verbal affirmation approach. A few researchers have created voice-command-based structures, such as wheelchair commands, for individuals with dysarthria [4].The findings of growing research on the employment of voice affirmation cerebrum linkages will be discussed in this work. This piece can be linked to the ID of verbalization, where the goal is to see someone speaking the words aloud.

| SNO | AUTHORS | ARTICLE TITLE | JOURNAL & VOL | OUTCOME |
|---|---|---|---|---|
| 1 | Pankaj Pathak et al | Speech Recognition Technology: Applications & Future | International Journal of Advanced Research in Computer Science, 1 (4), Nov. –Dec, 2010, 77-7, | This paper provides an effective approach focused on voice recognition to offer a simple control system. |
| 2 | Uma S*a ,R.Eswaria ,Bhuvanya Rb ,Gopisetty Sai Kumarb | IoT based Voice/Text Controlled Home Appliances | INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019, Uma S et al. / Procedia Computer Science 165 (2019) 232–238 | The users can merely provide voice commands or text messages through which they will be able to turn the appliances ON or OFF depending upon the necessity. The users can schedule the status of the appliances when they are not physically present in the environment. They will also b |
| 3 | R Ram kumar[1], A Vimal kumar[2], R Deepa[3], S Shalini[4] | VOICE COMMAND EXECUTION WITH SPEECH RECOGNITION AND SYNTHESIZER | International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 03 \| Mar-2018 | The speech recognition system development can be mainly used for speech recognition, speech generation, text editing, and tool for operating machine through voice. |
| 4 | Kristián Košťál , Pavol Helebrandt *, , Matej Belluš , Michal Ries and Ivan Kotuliak | Management and Monitoring of IoT Devices Using Blockchain † | Sensors 2019, 19, 856; doi:10.3390/s19040856 | The key feature of the blockchain, immutability, brings resistance to unauthorized modifications |
| 5 | Nwakanma Ifeanyi | text-to-speech synthesis | IJRIT International Journal of Research in Information Technology, Volume 2, Issue 5, May 2014, Pg: 154-163 | The goal of TTS research is to create natural-sounding speech that is indistinguishable from human speech. |

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 594-599 www.ijaem.net ISSN: 2395-5252

## III. METHODOLOGY

### EXISTING SYSTEM

Although some operating systems incorporate voice recognition technology [3] as a tool for decision-making, its full potential is not being completely utilized. Despite having a wide range of uses, its application is restricted to three stages. The recognizer must first be configured by the user. A sequence of random numbers will be shown when the user commands "show numbers" once the talk recognition engine has been deployed on the workspace. After that, the user chooses a number to carry out the intended action. This system's disadvantage is that it takes longer to execute. Speech recognition software used to rely largely on grammatical and linguistic rules.

The program could identify the spoken words if they followed a specific set of rules. Even when spoken consistently, human language includes a lot of deviations to its own laws. The way that particular words or phrases are spoken can be significantly impacted by accents, dialects, and peculiarities. The word "stable" might rhyme with "John" if someone from Boston were to say it without pronouncing the "r" The majority of people, on the other hand, do not pronounce their words clearly, like in the line "I will see the ocean." Third-party applications can be integrated with the current design, and the speech recognition technology can be utilized as a helper. Using a command like "show numbers" to access external programs will accomplish the desired result.

### PROPOSED SYSTEM

The suggested system aims to increase the effectiveness of speech recognition technology and make it possible to carry out extensive user voice commands. This is especially helpful in the fast-paced world of today, where people would rather give voice commands than use a graphical user interface (GUI) to do things [4]. In contrast to existing speech recognition systems, the suggested system uses a speech synthesizer algorithm that allows hands-free operation and reduces latency. With the assistance of a coach, the user can change the three command kinds that the system supports: social, web, and shell commands.

A speech synthesizer[5] is used to convert written text into spoken language, and the data message is processed to determine where sentences and other structures begin and end. The text is also examined for special forms of language, such as contractions, dates, times, numbers, and email addresses. Each word is subsequently transformed by the algorithm into a phoneme, which is a language's fundamental unit of sound. Lastly, each sentence's sound waveform is produced using the prosody and phonemes. The purpose of the suggested system is to give disabled people a voice-activated interface so they may use a computer without using their hands. This device is perfect for use in a manufacturing facility or while driving since it allows users to issue commands without the usage of a mouse or keyboard. Complex commands may be handled by the system with speed and efficiency, which improves performance and saves energy.

**Login**

In addition to providing essential information for other modules, the component allows the trainer to supply authentication details for verification. Similar to a notepad, the verification information are kept in a text file. Fields like location, username, Gmail ID, and password are all part of the login module. The file that is incorporated into the code contains the user's credentials. Only the trainer is subjected to the verification process.

**Synthesizer**

A voice synthesizer is used to translate written material into spoken language [12]. Another name for this procedure is Text-To-Speech (TTS) conversion [6]. Analyzing the input message to determine the beginning and ending of various sentences, segments, and other structures is the first step. This stage makes use of intonation and stress information in the majority of languages. After that, the text is analyzed to find any unique linguistic patterns.

For example, contractions, abbreviations, dates, times, numbers, currency amounts, email addresses, and other formats in English demand extra care. Each word is then broken down into phonemes, which are a language's basic units of sound. For instance, there are about 45 phonemes in US English, which include vowels and consonants. Lastly, the sound waveform for each sentence is created using the phonemes and information about intonation and stress. The phoneme and intonation information can be used in a variety of ways to generate the spoken output.

**Affix commands**

Three different command kinds are used by the system: Shell, Social, and Web commands. The Shell command delivers pertinent commands and first locates all files, folders, and applications. Informal language is difficult for the speech

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 594-599 www.ijaem.net ISSN: 2395-5252

recognizer to understand, but this module enables the inclusion of any application. The Web command respects the system's firewall and web security while permitting access to web pages using the built-in web browser. To respond to "what" questions, the question-response system uses the Social command.

**Directory based system**

The shell[8] request module is a significant feature of the system and operates on a file-based system. It manages the inventory of a specific file or action. The file serves as a means of accessing a document from the system. To access the shell request module, users typically log in to the system. This module can be used by the coach, who can update the list-based commands in the system with the help of the shell request module.

**Web based system**

The web request is a structure in the system that is used to access web addresses. This module can be used to add any collection of URLs to the system. Login credentials are required to access the web request module. Once the coach has logged into the system, they can access this module. Only with the coach's permission may the client use the updated requests. Since the client is usually regarded as disabled, the coach has access to update the requests.

**Request response system**

The "W" type of requests are handled by the system's Social request request-response module. It makes it possible for the system to engage with the user intelligently. Although social requests are an important kind of request, updating the data is a constant effort because user wants can change. The default queries that are shown to the user are predicated on the specifications that the user has given to the mentor. A login is necessary in order for the mentor to access the system's social request updates. One particularly clever module is the request-response module.

**System Architecture**

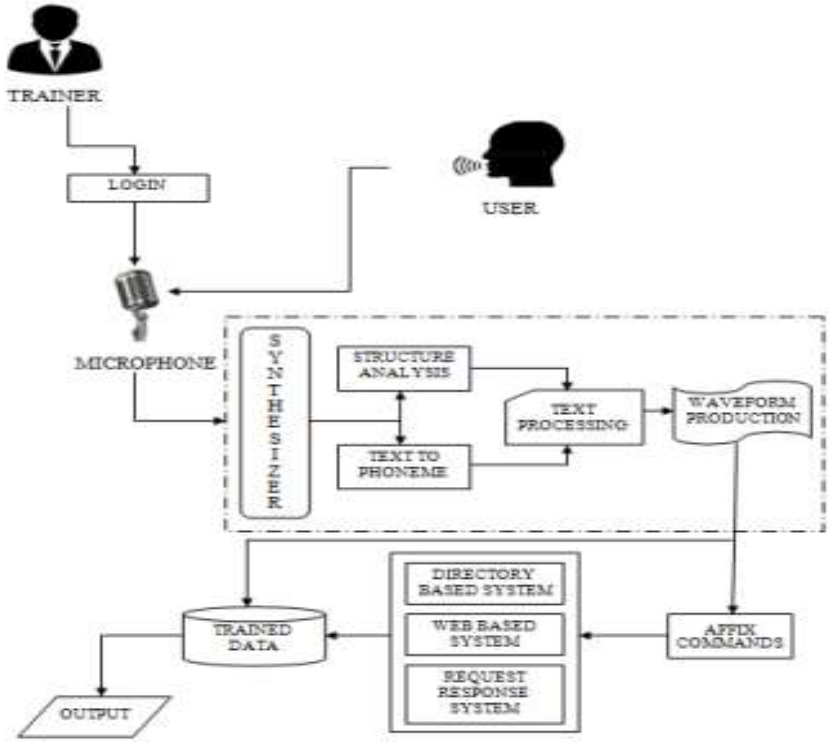


Fig -1: System Architecture

Creating a strategy or framework for the development of a system, service, or product is part of the setup process, which may involve the use of circuit diagrams, business process diagrams, or architectural drawings, among other diagram kinds. The process of transforming requirements into a model representation is known as setup in software engineering. Both text-to-speech and speech-to-text

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 594-599 www.ijaem.net ISSN: 2395-5252

conversions are possible with this program since it is made to recognize and produce speech. Through a microphone, the user gives vocal commands to operate the program. These commands are subsequently transformed into digital signals and analyzed as acoustic models. At runtime, the program executes dynamically, carrying out the specified program as required.

**ALGORITHM**

Text Pre-Handling Analyzing the data text for distinct linguistic forms is necessary. Abbreviations, contractions, dates, times, numbers, currency amounts, email addresses, and other formats of the English language require specific processing. The purpose of these extra procedures is to translate the written material into spoken language.

Text-To-Phoneme conversion The next stage is to break down each word into its phonemes, which are a language's fundamental units of sound. For example, there are roughly 45 phonemes in US English, which include both vowel and consonant sounds. The word "times" is pronounced as "t ay m s"—four phonemes, for example. Every language has a unique set of phonemes, or sounds. The digitized audio stream is then transformed into the fundamental waveform by the next steps.

Prosody AnalysisTo ascertain the proper prosody for the sentence, examine the words, phonetic elements, and sentence structure. Beyond the actual words being uttered, prosody includes a wide range of speech characteristics, including pitch (or melody), timing (or rhythm), pauses, speaking rate, word stress, and other elements.

**WAVEFORM PRODUCTION**

Finally, each sentence's sound waveform is conveyed through the phonemes and prosody information. From the phoneme and prosody information, the talk can be made in a variety of ways.

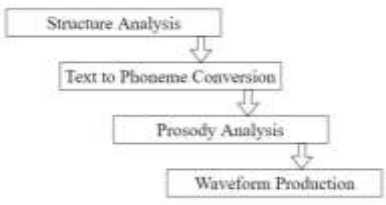


Fig -2: Synthesizer Structure

## IV. RESULTS AND OUTCOME

A set of Microsoft software tools known as Microsoft. NET[7] facilitates the rapid development and integration of Windows-based programs, Web designs, and XML Web services[8]. Programs can be created that can simply and securely interoperate thanks to the.NET Framework, a platform that is agnostic of programming languages. Java Script, Visual Basic, C#, and Managed C++ are among the languages that.NET supports. The.NET Framework serves as the foundation upon which ADO.NET is constructed. This Framework is an essential part of Windows that makes it easier to create and run new apps and XML Web services. Three main ADO.NET Data Providers are part of the.NET Framework.

These providers are designed to provide a code-execution environment that reduces programming interaction and eliminates formatting conflicts. SQL Server [10] uses the SQL Connection object, OLEDB uses the OLEDB Connection Object, and ODBC uses the ODBC Connection Object.

## V. CONCLUSION

The proposed system is an innovative approach to improve the efficiency of speech recognition technology and enable users to issue large-scale voice commands. By utilizing a speech synthesizer algorithm, the system enables hands-free operation, resulting in reduced time delay compared to current speech recognition systems. The system is designed to provide a voice-controlled interface for disabled individuals to operate a computer without the use of their hands. It supports three types of commands, including social, web, and shell commands, which can be updated by the user with the help of a coach. The system architecture includes a login component, a synthesizer module, an affix command module, a directory-based system, a web-based system, and a request-response system. The proposed system has the potential to improve performance and energy savings in industries like manufacturing, and it can also be used while driving. Overall, the proposed system is a significant step towards making voice recognition technology more accessible and efficient for us

## REFERENCES

[1]. D.Bandyopadhyay, Internet of things:Applications and challenges in Technology and standardization,

published online:9april 2011,Springer Science+Businessmedia,LLC.2011

[2]. Debasis Bandyopadhyay, JaydipSen, Internetof things: Applications and challenges in technology and Standardization

[3]. Pankaj Pathak et al, Speech Recognition Technology: Applications & Future, International Journal of Advanced Research in Computer Science, 1 (4), Nov. –Dec, 2010, 77-7,

[4]. L A Pratama and J Kustija 2020 IOP Conf. Ser.: Mater. Sci. Eng. 830 042053, Design of Graphical User Interface (GUI) on IoT-based remotelaboratory for Programmable Logic Controller (PLC) practicum andpneumatic simulation

[5]. Benoit C. Speech Synthesis: Present and Future. European Studies in Phonetic & Speech Communication, Netherlands. pp. 119-123(1995).

[6]. Nwakanma Ifeanyi,IJRIT International Journal of Research in Information Technology, Volume 2, Issue 5, May 2014, Pg: 154-163, text-to-speech synthesis (TTS)

[7]. gewarren. ".NET Framework & Windows OS versions". Docs.microsoft.com. Retrieved November 21, 2020.

[8]. Admin (3 August 2017). "What is a Web shell?".malware.expert. Archived from the original on 13 January 2019. Retrieved 20 December 2018.

[9]. "Web Services Architecture § Relationship to the World Wide Web and REST Architectures". W3C. Retrieved 11 November 2017.

[10]. "SQL Server 2008: Editions". Microsoft. Retrieved July 21, 2011.

[11]. Uma S*a ,R.Eswaria ,BhuvanyaRb ,Gopisetty Sai KumarbIoT based Voice/Text Controlled Home AppliancesINTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019, Uma S et al. / Procedia Computer Science 165 (2019) 232–238

[12]. R Ram kumar, A Vimal kumar, R Deepa, S Shalini,VOICE COMMAND EXECUTION WITH SPEECH RECOGNITION AND SYNTHESIZER International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 03 | Mar-2018

[13]. Kristián Košt'ál , Pavol Helebrandt *, , Matej Belluš , Michal Ries and Ivan Kotuliak Management and Monitoring of IoT Devices Using Blockchain † Sensors 2019, 19, 856; doi:10.3390/s19040856