AI Integrated Speech Synthesizer With Voice Automation

[1]Subi Eshawaran D,[2]Akshit Polepalli,[3]Simran Rani,[4]R. Logeshwari
[1]Student,[2]Student,[3]Student,[4]Assistant Professor
Dept. Computer Science Engineering
SRM Institue of Science and Technology
Chennai,India

***Abstract*—** Speech constitutes the primary form of human communication. It is capable of replacing keyboard, mouse and other peripheral devices. A speech recognition interface will be a beneficial innovation that supports majority of the essential applications, if established with an effective and efficient framework. It has many potential benefits and is useful to people in many ways. The one ultimate factor that differentiates humans from everything else is intelligence. Human beings comprehend things by figures, whereas machines understand things by instructions and knowledge. Knowledge Engineering is accomplished by understanding how human brain works and implementing the abstracted knowledge to the system. This knowledge includes, problem solving skills, decision making skills, response and reflex skills, memorizing skills, delivery skills, etc. The creation of an intelligence machine started with the intention of making machines with intelligence like human's brain. This concept is implemented here to have a potency to find and regard high in humans as their knowledge. So a speech recognition system, to be a successful one, it should contain abundant knowledge and dynamic response skill. This study will combine both speech recognition and artificial intelligence.

***Index Terms*-Artificial Intelligence, Voice, Speech Synthesizer, Voice Automation**

I. OVERVIEW

Speech recognition and its fundamentals can be understood with its operating procedures and the applications of speech recognition in varied square measures. [1]The speech recognition system is enforced as a desktop application within the given system. The speech recognition system development can be used to recognize speech, generate speech, redact speech, and act as a gear to operational systems by speech. The speech recognition system is additionally referred to as automatic speech recognition. The text-to-speech is implemented in the given system. Especially in english language, the words need exclusive analysis as is contains lots of abbreviations, acronyms, months, times, numbers, dates, currency, and lot more forms[1]. The speech recognition should be dynamic, in order that the execution times are reduced. Identifying the aptitude of speech recognition system and dealing to create its potency additional to execute high level voice commands given by the user. This finds great advantage in the moving world where the user has to give only voice commands to finish his job. Consequently this application is predicted to cut back the time delay in corporal punishment commands with GUI. Synthesizer algorithm is used in the proposed system. It is Hands-free computing. Reducing the time delay compared to the prevailing speech recognition system. User must offer solely voice commands to end his task. Three varieties of commands like social commands, web commands and shell commands are used in the proposed system. These commands will solely be updated victimization the trainer whereas user will solely use the commands. The trainer is attested employing a security level to access the change of commands into the system.

1.1 RELATED WORKS

A literature survey was conducted in the year 2017 for [7]a low-yield speech recognizer with a voice activity detector. It used DNN(deep neural networks) algorithm[7]. Previously, in the year 2013, another literature survey was conducted on [8]a speech recognition system in noise free environment[8]. Blind Source Extraction (BSE) is an attractive approach to enhance multichannel noisy speech data, which acts are a prerequisite for speech recognition system.

II. SYSTEM OVERVIEW

The system components and the processing of components are elaborated. Design Engineering deals with the various diagrams for the implementation of project. This software is crafted to analyse the speech and also has the ability to speak and synthesize. [12]It is capable of transforming speech to text and text to speech. The user has to feed voice command as the input data via microphone and microphone obtains the command and the analog signals are transformed to digital ones in the internal circuit. These digitized signals are processed as acoustic model. The windows grammar verifies the command as a valid one in its default language. Normally the language used in the system is Standard English language[12]. Then the speech recognition model comes into act. The speech recognizer application in windows 8 is connected through .NET in visual basics where the operational code is written in C# and Visual Basic invokes the application in front end. The commands are stored in the inbuilt file present as an embedded form into the code. Once the command is identified the application contemplates the command with the inbuilt code to execute the corresponding function. The program is essentially executed at run-time as the given program is dynamic. The architecture is given in fig.1

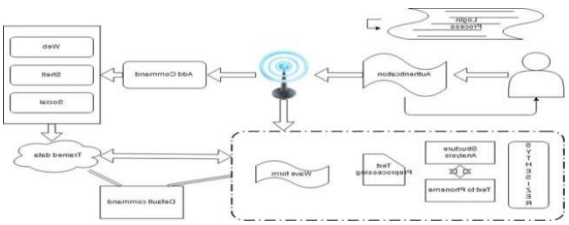

Fig 1.Architecture

II. REQUIREMENTS

2.1 HARDWARE REQUIREMENTS

The application is capable of providing service with high efficiency. Hardware requirements include resource requirements and other essential components to run the application at an optimum level. The most common requirement demanded by any application is the physical hardware resources. Hardware requirement set is usually listed in the hardware compatibility section. The following sections discuss the assorted aspects of hardware needs. The hardware requirements constitutes the hardware resource requirements and pre-installed essential hardware components that are needed to be plugged-in to a system to deliver optimal processing of the application. The hardware requirements are as follows, dual core processor 2.00GHZ, 140GB internal storage, microphone, 2GB of RAM.

2.2 SOFTWARE REQUIREMENT

The needs which specify the software resource requirements and prerequisite softwares that are needed to be established to generate finest function of the application are called software requirements. Software requirement set is usually listed in the software compatibility section of the software description. The software requirements are as follows, windows7 and above, ms visual studio .net 2010, ADO.NET and C#.

III. SYSTEM IMPLEMENTATION

3.1 SOFTWARE DESCRIPTION

3.1.1 DOTNET

[2]MS .NET framework is a library of software packages used to develop applications such as XML service application, MS windows applications and other internet solutions. The DOTNET Framework is a language neutral platform which provides a platform for developing programs which will simply and firmly inter-operate. DOTNET provides language freedom due to it's widespread language compatibility environment. DOTNET applications can we written in C#, F#, visual basic, etc[2]. The .NET framework delivers the ability for elements to move smoothly, regardless of the platforms.

3.1.2 ADO.NET ARCHITECTURE

It is completely based on fundamental architecture of the DOTNET Framework. MS visual studio provides properties to mix the varied options of this design. The .Net Framework includes chiefly 3 knowledge suppliers for .NET. .NET architecture is given in fig.2

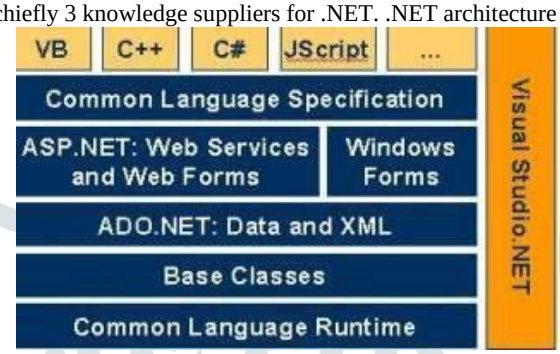

Fig 2. .NET Architecture

3.1.3 MICROPHONE USAGE

Microphones use magnetic force induction, capacitance amendment or piezoelectric effect to provide associate degree motorized signals from atmospheric differentiation. [11] Microphones are connected to a component called pre-amplifier. Pre-amplifier is a component which prepares the frequencies for amplification and rendering. The rendered frequencies are sent for amplification to get effective range of frequencies. The microphone determines and analyses the speaker's input dynamically in the direction of frequency. This creates an electromagnetic radiation annd reverb from alternative elements of the area doesn't seem to be picked up[11]. In addition, digital noise reduction process removes background.

3.1.4 COMPONENTS OF MICROPHONE

[3]The sensitive electrical device part of a mike is named as its part. Except the layer primarily based electro-acoustic transducer, sound is 1st regenerated by a diaphragm to mechanical motion, the locomotion of that is then rejuvenated to Associate in Nursing electronic communication. A veritable mike which includes a housing additionally, some suggests that of transferrable the indicator from the part to different instrumentation, sometimes Associate in Nursing electronic periphery to accept the output data of the capsule to the device which is being driven[3]. A broadcast transmitter includes a tuner microphone.

3.1.6 SPEECH SYNTHESIZER

The transcription is converted into auditory communication by speech synthesizer. It is also referred to "Read Aloud" technology ,which can be also called initial speech recognizer.

3.1.7 ANALYSIS OF THE STRUCTURE

Processing to work out the input text where the paragraphs, sentences and alternative structures begin and finish. For most of the languages, punctuation and format information area unit utilized in this stage.

3.1.8 PRE-PROCESSING WRITTEN WORK

For the noteworthy build of the language, the input text is examined. For the ussage of diminutive, shortening, dates, times, figure, coinage, E-message and plenty of alternative mould special treatment is needed in English language. Other dialects requires exceptional process for these conformation and most of the vernacular produce authoritative necessities. Transformation of the spoken text to speech is as follows

3.1.9 TEXT TO ALLOPHONE CONVERSION

[4]Converting each anaphor to speech sound. In every language, a syllabary could be a phonology. US English contains in and around forty five hieroglyphs together with the consonant and operatic sounds[4]. For example, "apple" is oral as two characters "ap el". Different languages have totally different sets of sounds (different morphemes). For exemplar, Japanese has few signs together with resonate which are not found in Received Pronunciation, like "its" in "tsunami". The remaining steps convert the digitalized voice signal to the analog wave shape.

3.1.10 PRASODY ANALYSIS

[8]Processing to work out applicable poetics for the pronouncement that has structure, words and rune. Prosody has several of the options of speech. This covers the tone (or melody), the scheduling (or rhythm), the halting, the talking rate, the prominence on words and many other trademarks[8].

3.1.11 WAVE FORM MANUFACTURE

[8]Finally, the morphemes and metrics data turn out the aural undulation as much as every verdict. There are some methods within which the verbal communication is often made from the articulation sound and poetic meter data[8].

The overall flow of synthesizer is given in fig. 3

Fig 3.Waveform production

3.2 LIST OF MODULES

3.2.1 LOGIN

The module allows the trainer to provide the trainer information for authentication. It provides all basic information for other modules. [9]The authentication details are stored in text file (i.e.) notepad. The contents of the login module are location, username, Gmail ID and password. The user given credentials are stored onto the file embedded to the code[9]. The authentication is only done for the trainer.

Architecture of login module is given in fig.4

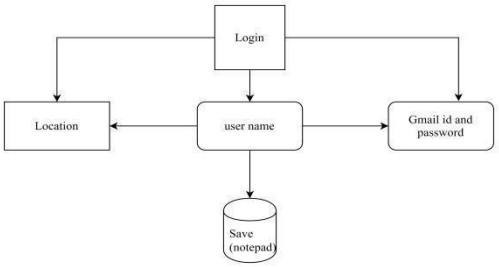

Fig 4. Login module

3.2.2 ADD COMMANDS

There are three types of commands used in the system. They are Shell command, Social command, and Web command. [10]The shell command stores Location of each file, folder and application is to be specific at the start to trainer. It is recommended to provide related commands. Colloquial languages are difficult to recognize for speech recognizer. Any application integration is finished the assistance of this module. Using web command, web pages can be accessed using your default web browser with the help of this module. The social command is employed for request-response system that is employed for "what" form of queries[10]. Add command module architecture is given in fig.5

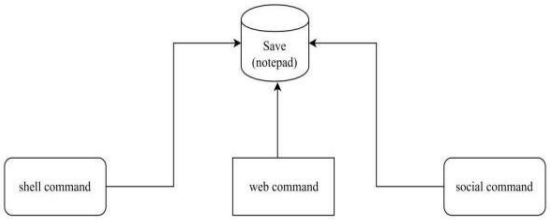

Fig 5. Add commands module

3.2.3 SYNTHESIZER

[5]A speech synthesizer changes transcription to language. It is also called as text 2 speech conversion. The text which is given as input is processed to see wherever paragraphs, sentences and alternative structures start and end. Punctuation and data formatting knowledge square measure is done in this stage. The input data is thoroughly analyzed for any constrains of the semantics and syntax. Each and every word is converted into phonemes to derive the frequency of it[5]. United states English has around forty five phonemes together. Finally, the phonemes and prosody data square measure accustomed manufacture the audio wave shape for every sentence. There square measure some ways during which the speech may be made from the phone and prosody data. Synthesizer module architecture is given in fig.6

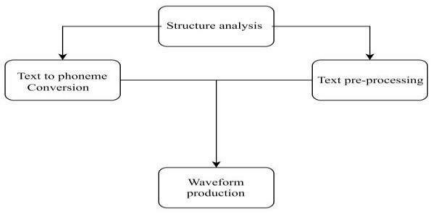

Fig 6. Synthesizer module

3.2.4 SHELL COMMAND

Shell command is one of major enhancement made in the system. The shell command deals with the directory of any explicit file or action. The directory may be a path or thanks to access any form of file from the system. The trainer will solely update the directory based mostly commands into the system with the utilization of shell module. Shell command module architecture is given in fig.7

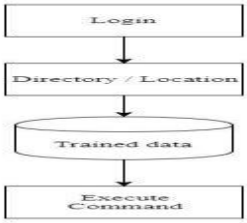

Fig 7. Shell commands module

3.2.5 WEB COMMANDS

The internet command is that the web primarily based command system. It is increased to primarily access the Uniform Resource locater (URL) within the network. Any set of Uniform Resource locater (URL) is supplemental to the system. The web module has some permission to be glad to access it. The login should be done before accessing the net module. As the trainer will solely login the system, the net module will solely be accessed by the trainer. The user has the permission to use the updated command by the trainer. Mostly user is taken into account as individual; thence trainer is given the access to update commands. Web commands module architecture is given in fig.8

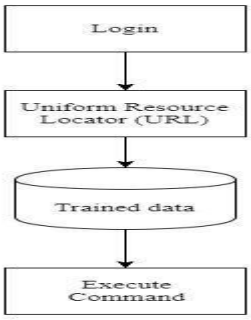

Fig 8. Web commands module

3.2.6 SOCIAL COMMAND

Social command is that the request response system, that is that the "W" variety of question asked to the system. Social command is the vast type of command whereas updating of data is a continuous process, as the requirements of the user may vary. The questions framed for the user as default are obtained from the user given requirements to the trainer. The trainer has the credibility to access the updating of social command into the system, as the login can only done by the trainer. Social command module architecture

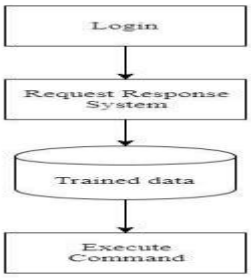

Fig 9. Social commands module

IV.EXPERIMENTAL RESULTS

Identifying the potential of speech recognition system and dealing to create its potency additional to execute high level voice commands given by the user. This finds great advantage in the moving world where the user has to give only voice commands to finish his job. Consequently this application is predicted to cut back the time delay in capital punishment commands with interface. Synthesizer algorithm is used in the proposed system. It is Hands-free computing. Reducing the time delay compared to the prevailing speech recognition system. User has got to offer solely voice commands to complete his task. Three varieties of commands like social commands, web commands and shell commands are used in the proposed system. These commands will solely be updated victimization the trainer whereas user will solely use the commands. This speech synthesizer is designed by keeping handicapped individuals. It will pave a new path for them by making them operate computer with ease. Hands-free computing may be a program during which user will work while not the employment of hands, a standard demand of peripheral units like keyboard and mouse are made secondary in the proposed system. Speech synthesizer can be trained to receive commands, and once the command is confirmed by the user, instructions will be fed into the system without the usage of keyboard or mouse. The disabled will be benefited by this hands- free experience of computing. Speech recognition system can be trained. The planned system will answer complicated queries or task given by the user because it performs the action in faster time delay. Energy is saved efficiently as the performance of the system is increased.

4.1 OUTPUT

Following fig.10 is the welcome window which appears when the application is started. It consists of three sections, first section is the title of the application, second section is the welcome note, third section is the enter button which when clicked starts the processing of the application.

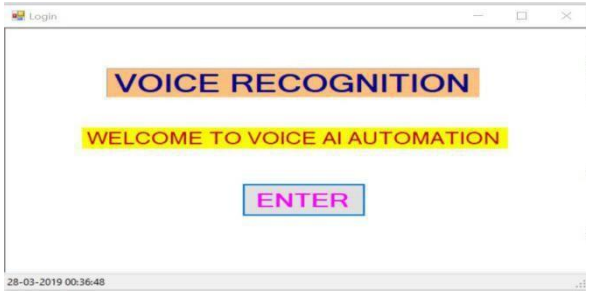

Fig 10. Welcome window

Following fig.11 is the operational window which appears when the user clicks the enter button. The voice synthesizer notifies the user that it is ready to take voice commands.

Fig 11.Operational Window

Following fig.12 is the admin authorization window. It consists of two input sections, "USERNAME" and "PASSWORD". By entering the following credentials, the user will be granted administrator access. An administrator can add, delete and modify commands.

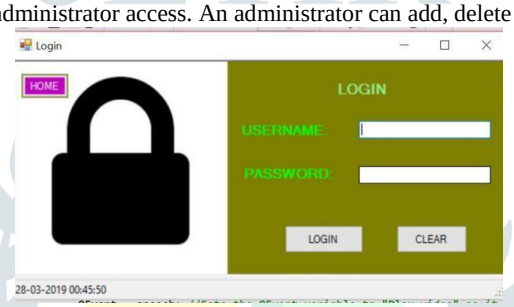

Fig 12.Login Window

Following fig.13 is the administrator workspace. In here, the admin can add, delete or modify commands. There are four sections in this window, shell commands, web commands, social commands, email and weather.

Fig 13.Admin Workspace Window

Following fig.14 is the shell command window. Under this section, the admin can add, remove or modify the system commands. The internal processes are taken care by this shell commands section.

Fig 13.Shell Command Window

Following fig.14 is the web command window. Under this section the admin will be able to add commands which can access the internet. For example, if the user trains the application to open google.com, it will access www.google.com when the voice command is given.

Fig 14.Web Command Window

Following fig.15 is the social command window. In this section, the admin will be able to add general conversing commands. For example, if the admin trains the application to respond to a "good morning" greeting, it will respond when the voice message is given.

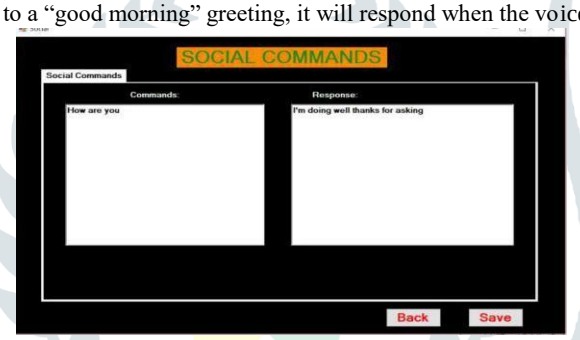

Fig 15.Social Command Window

Following fig.16 is the email and weather window. This section requires the email login credentials. This part of the application gets access to the email database and the weather updates.

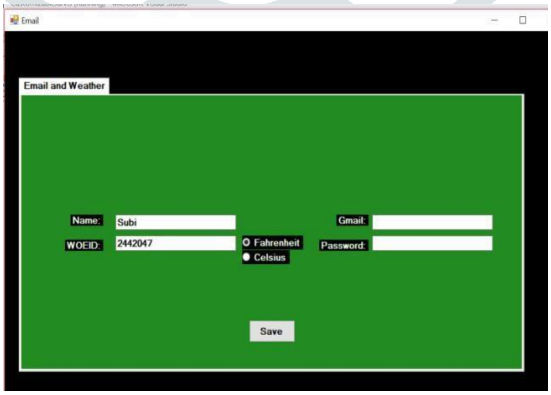

Fig 16.Email and Weather window

V. CONCLUTION

A speech recognition system called voice synthesizer was proposed and the relative process was described briefly. This project supports computer code development for speech recognition. In early stages, the application was processed with minimal data and tools. At the later stage, we have implemented completely different tools to deliver a sensible work as output. The computer code has been debugged and tested and results were mentioned. Improvement and modifications of the modules are done in the final stage of the development and the same will be done in the future if needed. This application is built keeping certain key requirements in mind. Those requirements are fulfilled and the software is ready application in real computing world.

VI. REFERENCES

[1] URL:https://en.wikipedia.org/wiki/Speech-generating_de vice.
[2] .NET URL: https://en.wikipedia.org/wiki/.NET_Framework.
[3] Microphone. URL:https://www.maplesoft.com/support/help/maple/view.aspx?p ath=MicrophoneComponent
[4] M.H. O'Malley."Text-to-speech conversion technology", IEEE Journal Aug. 1990.
[5] Youcef Tabet;Mohamed Boughazi. "Speech synthesis techniques. A survey. 27 June 2011.
[6] Chandrakasan. "A Low-Power Speech Recognizer and Voice Activity Detector Using Deep Neural Networks" , IEEE Journal of SolidState Circuits,2018
[7] Suma Swamy and K.V Ramakrishnan computer science & engineering: An International Journal (CSEIJ), vol. 3, no. 4.
[8] M.B. Yeary;R.J. Fink;D. Beck;D.W. Guidry;M. Burns."A DSP-based mixed-signal wavefrm genrator".18 May 2004
[9] Authentication Module. URL:https://codingcyber.org/simple-login-script-php-andmysql-64/.
[10] Command module. URL:https://hackernoon.com/how-to-train-your-robot-ai-fo r-everyone-69b96ad943e5.
[11] J.S. Park, G.J. Jang, J. H. Kim and S.H. Kim, "Acustic Intrference Cancellation fora Voice-drivn Interface in Smart TVs."
[12] Suzuki, Dept. of Comp. Sci., Nagoya Inst. of Technol., japan zen, h. ; Nankaku, Y. ; Miyajima, C. ; Tokuda, K. ; Kitamura."Titlespeech recognition using voice-characteristic-dependent acoustic models". Oct 2011.