DNN AND DL ALGORITHM CONVERSION OF HUMAN VOICE DATA TO ENGLISH TEXT

[1]**Dr. N. Chandra Mouli,** [2]**Y.Susheela,** [3]**D. Sai shravani**

[1]Associate Professor, [2]Assistant Professor, [3]Student [1,2,3] Dept. of Computer Science Engineering, [1,2,3]Vaageswari College of Engineering, Karimnagar, Telangana
E-Mail: cmnarsingoju@gmail.com, chiduralasusheela@gmail.com

ABSTRACT

Speech is the most acceptable form of communication amongst human races. Human-computer interface refers to the interaction between humans and computers. Speech recognition translates spoken words and phrases into a machine-readable format after recognizing them. Speech recognition is a technological advancement that identifies words and expressions on its own that most closely match the value of human speech. This development enables the development of a broad range of applications in which the output of speech recognition can either be the final product (such as letters) or a contribution to further preparation (e.g., providing orders to gadgets dependent on the identified words, semantic examination of discourse in a programmed exchange framework). The most natural use of Speech Recognition is converting our Speech to Text for various Purposes and also for using it as a tool to control our Computers

Keyword: Speech, Text, DNN, DL.

1. INTRODUCTION

Inside the path of recent years, cellular phones have emerged as an irreplaceable wellspring of correspondence for the cutting aspect society. We are able to come to a decision choices and instant messages from a supply to an objective without any trouble. Its miles realized that verbal correspondence is the most suitable modem of passing on and imagining the proper facts, avoiding misquotations. Speech reputation which is likewise known as automatic speech recognition (ASR) and voice recognition acknowledges the spoken phrases and phrases and converts them to gadget-readable layout via converting spoken audio into text, speech recognition technology permit customers to control virtual gadgets by way of talking in place of the use of conventional gear consisting of keystrokes, buttons, keyboards and many others. Controlling things using voice command has constantly been an obsession for humans. In recent times we see several voice operated appliances all around us. Those devices are capable of doing many things like controlling appliances. But, all of these structures are pre-configured and it's miles quite hard to do any modifications in them. Also, those services are on a dedicated tool or else a telephone. Those styles of device are not available on laptop computer systems and laptops. Here, our idea is to increase a python based totally open source speech to text engine that have a high degree of customizability and may be deployed without problems of several working systems. The gadget will allow person to carry out more than one obligations primarily based on voice controlling best.

2. RELATED WORK

Presently, existing are simplest available on devoted platforms and smartphones. Such offerings haven't been implemented on computer systems well. A person should do the whole thing using traditional method via Keyboard and Mouse. Proposed gadget is an open source very interactive voice-primarily based device with user pleasant response developed with the usage of Python Programming language. The machine will allow consumer to do more than one duties like open an software, typing and several other matters entirely based on voice instructions. The developed gadget can run easily on numerous working systems like windows, MacOS, Linux

and so forth. Additionally, the machine will provide a excessive degree of customizability. Also as a multi-language help, the system could be able to knowledge essential languages that's English and Hindi.
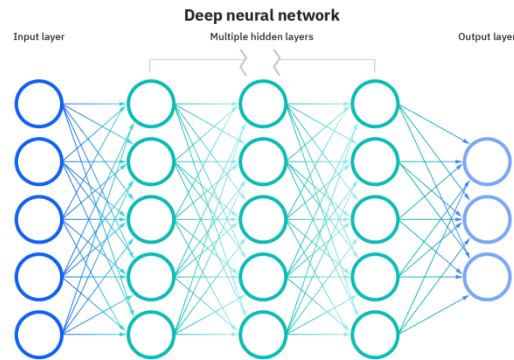
3. IMPLEMENTATION

DNN

Fig.1. deep neural network

A deep neural network (DNN) is an artificial neural community (ANN) with more than one layer between the input and output layers.[2] There are specific types of neural networks, however they constantly consist of the identical components: neurons, synapses, weights, biases, and capabilities. These components functioning similar to the human brains and can be trained like another ML algorithm. For example, a DNN that is skilled to understand dog breeds will move over the given photo and calculate the probability that the dog within the image is a certain breed. The user can evaluate the consequences and pick which chances the network should show (above a positive threshold, etc.) and go back the proposed label. Each mathematical manipulation as such is considered a layer, and complicated DNN have many layers, consequently called as "deep" networks. DNNs can version complicated non-linear relationships. DNN architectures generate compositional models in which the object is expressed as a layered composition of primitives. The extra layers allow composition of capabilities from decreased layers, potentially modeling complex facts with fewer gadgets than a similarly appearing shallow network. As an instance, it became proved that sparse multivariate polynomials are exponentially less difficult to approximate with DNNs than with shallow networks. Deep architectures consist of many editions of a few primary processes. Every architecture has located fulfillment in particular domain names. It is not usually possible to compare the overall performance is more than one architecture, until they had been evaluated at the same data units. DNNs are generally feedforward networks wherein data flows from the input layer to the output layer without looping again. At the beginning, the DNN creates a map of virtual neurons and assigns random numerical values, or "weights", to connections among them. The weights and inputs are improved and go back to an output between zero and 1. If the network did not correctly understand a selected pattern, an algorithm could modify the weights. In that manner the algorithm can make certain parameters extra influential, till it determines the best mathematical manipulation to completely procedure the facts.
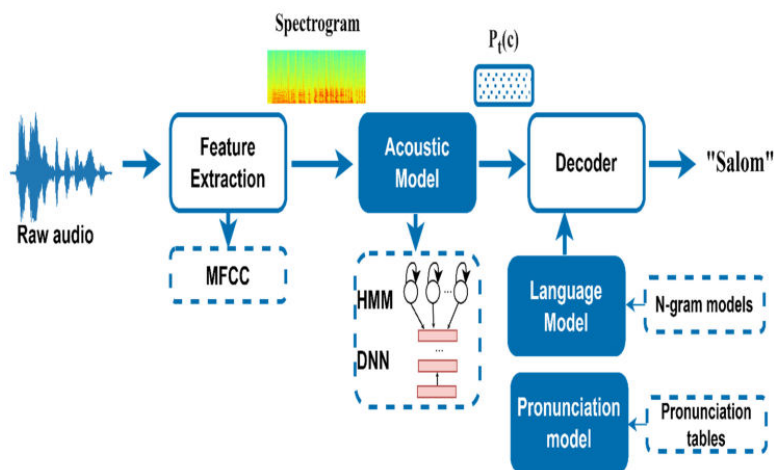
Fig.2. DNN also have been applied to acoustic modeling for automatic speech recognition (ASR).
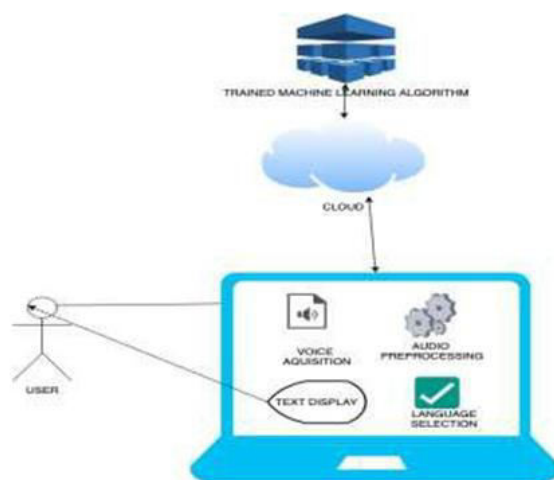

Fig.3. System Architecture

Modules are files that contain Python definitions and declarations. Modules can define functions, classes, and variables. Modules can also include executable code. Grouping related code in the module can make it easier to understand and use the code. It also makes the code logical. In Python programming, treat modules as the same as code libraries.

Language Selector Interface Module

This module will be for showing an UI to user for selecting the appropriate Language for Voice to Text Conversion.

Audio Recording Module

This module will be used for getting access to the Computer's Microphone and recording the User's voice and storing it in a temporary file.

Audio Pre-processing Module

This module will be used for pre-processing the recorded audio. This will be used for features like removing whitespaces in audio, Cancelling or minimizing the background noise, detecting when the user stopped speaking, etc.

Speech to Text Conversion Module

This module will be responsible for sending the pre-processed audio to the cloud server for conversion of voice to Text and then receive the response from the Server.

Display Text Module

Once the Voice is converted into Text, This module will print the Text output on the Screen. This module will be also responsible for showing any error if occurs.
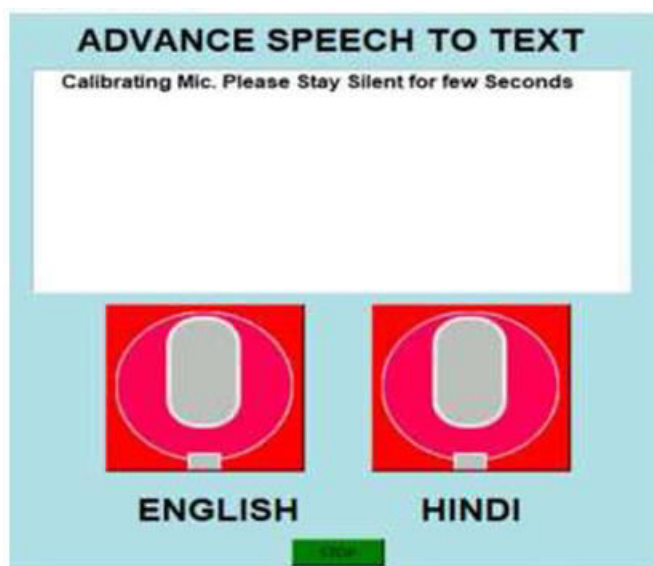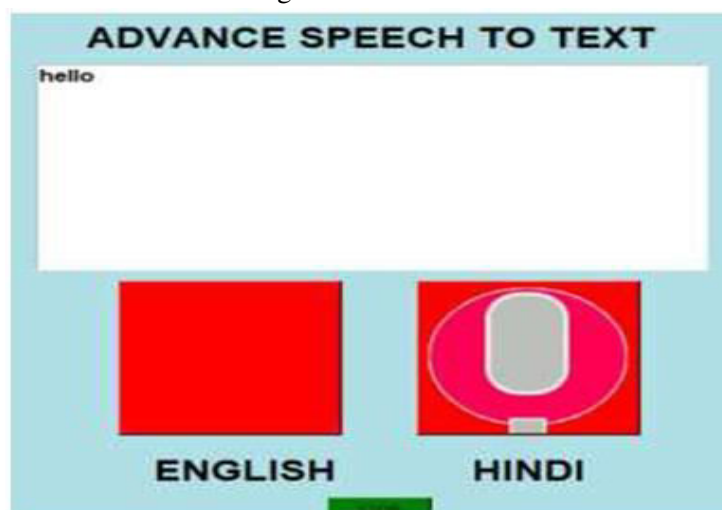
4. EXPERIMENTAL RESULT

Fig.4. Home screen

Fig.5. Result screen

CONCLUSION

Speech recognition was implemented, and the final product became a functioning system. The development of a speech-to-text system that could understand many languages and take commands to activate other software programmers will be adequately advanced and assessed. The device should accomplish the assignment's aim to the development of speech-to-text devices that can interpret a number of languages and take commands to launch various software applications. The device functions well and completes the mission's goal. A suitable

requirement was decided after analysis of the speech recognition. The device must accomplish the challenge's goal.

FUTURE ENHANCHEMENTS

The System can have more advanced feature in Future like: More Control over system from speech command in addition of more languages. Speech recognition may become speech understanding.

REFERENCES

1. Lemmetty, S., 1999. Review of Speech Syn1thesis Technology. Masters Dissertation, Helsinki University Of Technology.
2. Dutoit, T., 1993. High quality text-to-speech synthesis of the French language. Doctoral dissertation, Faculte Polytechnique de Mons.
3. Suendermann, D., Höge, H., and Black, A., 2010. Challenges in Speech Synthesis. Chen, F., Jokinen, K., (eds.), Speech Technology, Springer Science + Business Media LLC.
4. Allen, J., Hunnicutt, M. S., Klatt D., 1987. From Text to Speech: The MITalk system. Cambridge University Press.
5. Rubin, P., Baer, T., and Mermelstein, P., 1981. An articulatory synthesizer for perceptual research. Journal of the Acoustical Society of America 70: 321–328.
6. van Santen, J.P.H., Sproat, R. W., Olive, J.P., and Hirschberg, J., 1997. Progress in Speech Synthesis. Springer.
7. van Santen, J.P.H., 1994. Assignment of segmental duration in text-to-speech synthesis. Computer Speech & Language, Volume 8, Issue 2, Pages 95–128
8. Wasala, A., Weerasinghe R. , and Gamage, K., 2006, Sinhala Grapheme-to-Phoneme Conversion and Rules for Schwaepenthesis. Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions, Sydney, Australia, pp. 890-897.
9. Lamel, L.F., Gauvain, J.L., Prouts, B., Bouhier, C., and Boesch, R., 1993. Generation and Synthesis of Broadcast Messages, Proceedings ESCA-NATO Workshop and Applications of Speech Technology.
10. van Truc, T., Le Quang, P., van Thuyen, V., Hieu, L.T., Tuan, N.M., and Hung P.D., 2013. Vietnamese Synthesis System, Capstone Project Document, FPT UNIVERSITY