

Speech Recognition System in Regional Language based on AI and Machine Learning

O S KOUSHIK¹, T M SRINIVASA², Dr. SUMA S³

¹PG Student, Dept. of MCA, DSCE, Bangalore

²Assistant. Prof., Dept. of Computer Science and Engineering NIT Calicut, Kerala

³Associate Prof., Dept. of MCA, DSCE, Bangalore

Abstract - The most emerging technologies that are being used all over the globe today for real-time applications are Machine learning and artificial intelligence. Machine learning has many potential applications, including chatbots, data processing, industrial automation, and autonomous transport vehicles, amongst many others. On the contrary, the language processing system helps in voice recognition. It makes the device more versatile and reliable in terms of the background sound, languages, and sound frequency in the area where the user is. Consequently, it recognizes and responds to the regional language. When presented with abstract data, a computer system is trained via machine learning to make accurate predictions from the given data. Using methods from Natural Language Processing and Speech Recognition, the central theme is to design an automated system capable of recognizing the voices of speakers of regional languages.

Keywords: Natural Language Processing system, Hidden Markov Models, Artificial Neural Networks, Regional Language, Machine Learning, Machine Learning, Database.

1. INTRODUCTION

Machine learning (ML) is a method of data analysis that automates the construction of rational models. It's a sort of artificial intelligence since it allows systems to learn from data, see patterns, and make decisions with little human involvement. Artificial Intelligence (AI) technology is essential nowadays because it helps businesses get insight into customer behavior patterns and company functional examples, which in turn aids in developing brand-new goods. Automatic Voice Recognition (ASR), often called computer speech recognition or speech-to-text, is a function that enables computer software to transcribe human speech into text. Voice recognition refers to a device's ability to understand and carry out a user's spoken instructions. Thus, voice recognition enables hands-free control of various instruments and equipment, feeds automatic translation, and generates regional language output. Phonemes are the fundamental speech units, and a voice recognition system

may distinguish them by analyzing the digital input. The interprets human speech is known as Natural Language Processing (NLP). It also incorporates elements of AI. Using Natural Language Toolkits (NLTKs) mainly translates and automates the task corresponding to the user's input.

1.1 Voice Recognition / Speech Recognition in Machine Learning

There are many most popular Machine Learning (ML) techniques, such as Hidden Markov Model (HMM), Discriminative Learning (DL), Structured Sequence Learning (SSL), and Bayesian Learning (BL), and Adaptive Learning (AL). ML techniques are examples of methods typically motivated by ASR. Even more so, ML may use ASR as a large-scale^[1] practical application to evaluate the performance of a strategy rigorously and to create new problems stemming from the inherently sequential and dynamic nature of speech^[11]. However, even though ASR is commercially accessible for specific applications, the issue is still not resolved for the most part^[13]. On the other hand, the novel perspectives that contemporary machine learning approaches can hold a great deal of promise for the development of ASR technology.

1.2 Speech Recognition / Voice Recognition Algorithms

The HMM is a keyword recognizer based on a continuous speech recognition model and is not reliant on the speaker. In this section, we cover the primary keyword identification approach, non-keyword speech, and the implications of linear channels. Additionally, the linear channel effects, on the other hand, are amenable to precise correction by using a spectral normalizing procedure dependent on the acoustic class^[2]. In addition, the scoring technique is based on the likelihood ratio to consider the many elements that can alter the likelihood scores of terms^[3]. Initially, the phrase length will vary from one application to another even though it is the same term. Varying phrase length issues may fix by

simply normalizing the templates and the unknown^[4]. On the other hand, their lengths are the same, and there was no way to keep a steady pace of speaking throughout the whole thing. It is feasible that a nonlinear function of time will prove to be the most effective method for matching a template with a voice sample. The Dynamic Time Wrapping (DTW) techniques are valuable tools for determining the best nonlinear alignment and processing action recognition. Artificial Neural Networks (ANNs) train the user's input voice^[12]. Subsequently, recognized by speech recognition and been translated through NLP.

2. NLP WITH NATURAL LANGUAGE TRANSLATOR (NLT) TOOLKIT

NLP aims to allow computers to read and write as people do^[5]. Early computational methods for studying languages primarily focused on automating studies of linguistic structure and establishing essential technologies such as translation software, voice search, and speech recognition^[6]. Conversation systems, spoken translation engines, health and financial information mining on social networks, and attitude and mood recognition are examples of real-world applications that use similar approaches, which are currently being developed and put into use. Also expected is that machines will carry out a significant percentage of the translation process. Place the many approaches to machine translation^[8] into one of these three basic categories rule-based or knowledge-based. Considering the empirical evidence, the term "approaches" refers to a wide variety of methods that includes, but are not limited to, quantitative analysis and contextual inquiry, amongst others (interlingua, transfer, etc.)

This voice recognition software bundle has a wide variety of specialized access levels. The recognizer^[7] assists the system in interpreting the user's corresponding input by using a wide variety of Indian languages as examples during training.

3. MOTIVATION

Machine Learning and Speech Recognition System mainly signifies natural language processing, speech recognition, and automation. of languages around the world for natural language translation. The languages include just English, Chinese, Japanese, and Korean (NLT). Even though it is anticipated that the percentage of people living in India who own smartphones will increase from 66.21 percent to 95.81 percent by 2040^[14], no significant innovations or implementations have been made to emphasize translation into Indian regional languages. It is possible to save substantial time by having a voice recognition system that automatically recognizes, interprets, and responds to user input in a regional language. This is why a voice recognition

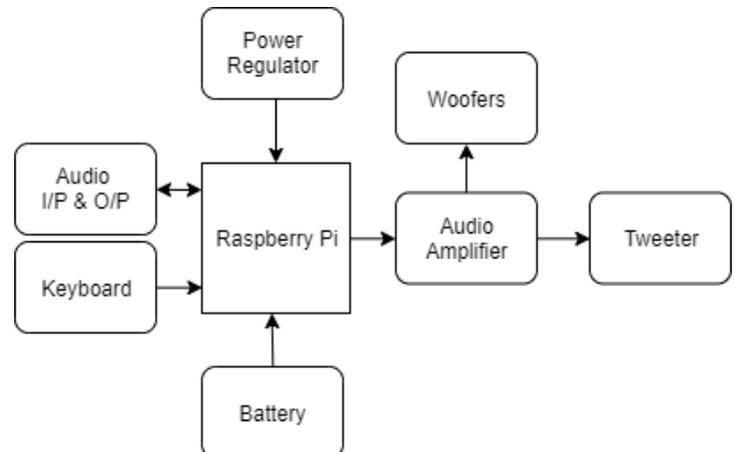
system based on machine learning and artificial intelligence is so important.

4. PROPOSED METHOD FOR VOICE / SPEECH RECOGNITION IN REGIONAL LANGUAGE

The user's voice is the primary input source for the machine learning and artificial intelligence-based speech recognition system that uses regional language. The system then identifies and translates the speech using NLP and NLT. Responds in kind to the user's commands and is therefore fully automated. Following that, the information is saved in a database.

4.1 Design of the Hardware

In a cylindrical tube, we place the woofers connected to the audio amplifier of TPA3116D2, which is connected to the tweeter and helps produce high-quality sound. Subsequently, Connect the power regulator to the audio amplifier. Now translate the Speech recognition into the User's input language concerning the NLTK and Automates shown in the figure below.



4.2 Software Interfaces

Python's flexibility and library support make it an excellent choice for building specialized machine-learning tools to address challenging problems like financial time series analysis, data analytics, etc. The Raspberry Pi is compatible with many operating systems. However, it is a more suitable tool for creating and implementing IoT, speech recognition, etc., since it has official support. As a result, we use the device mentioned above in Voice Recognition for regional languages.

4.3 PYTHON PACKAGES REQUIRED FOR SPEECH / VOICE RECOGNITION IN REGIONAL LANGUAGE FROM PYTHON

We must import the python packages such as **PyAudio**, **Speech Recognition**, **Google-trans**, and **NLTK**.

PyAudio, Port Audio, is a cross-platform C++ library interfacing with audio drivers.

Speech Recognition is a library that mainly recognizes the input of the user's voice.

Google-trans is a free application interface tool for translation.

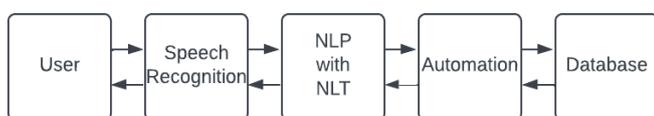
NLTK is a toolkit that makes translates.

5. ALGORITHM

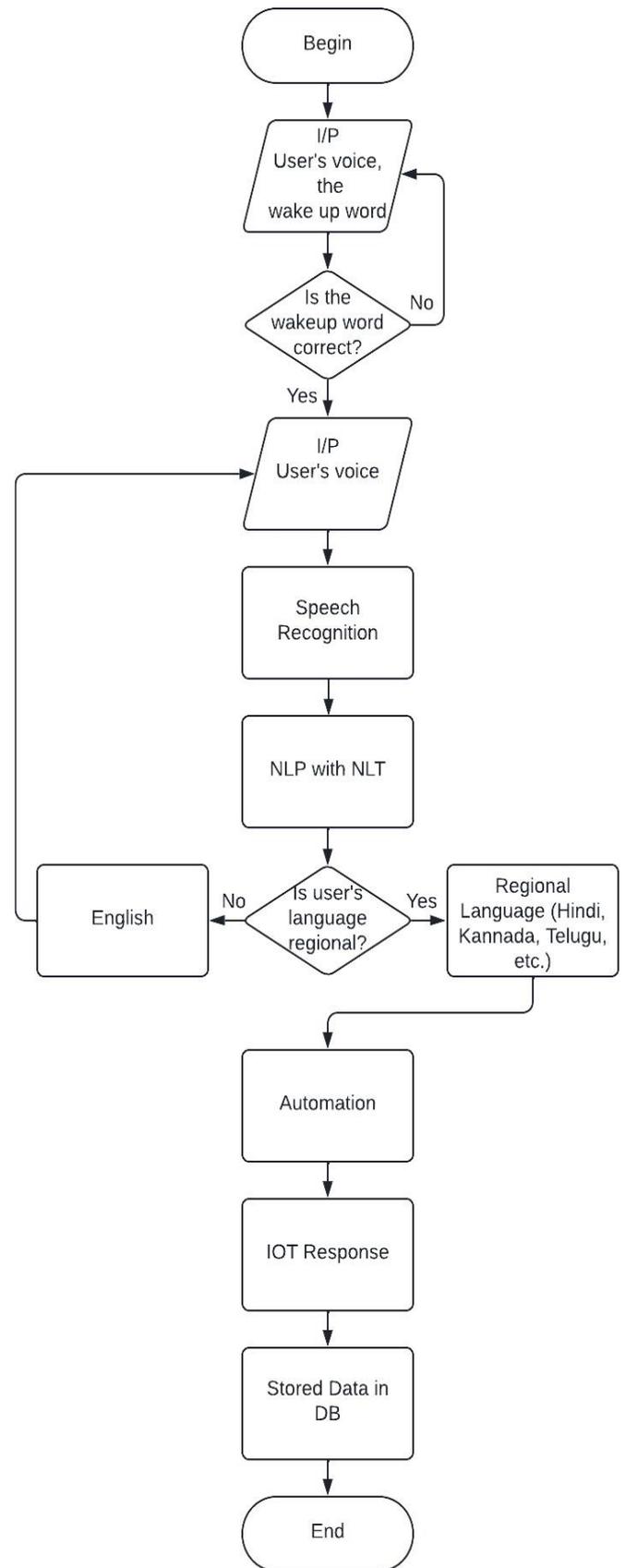
1. Input the natural language voice/speech as a command from the user to wake up the device.
2. The device responds if the input command is correct; otherwise, go to step 1.
3. Accept the user's input, i.e., voice/speech. For example: Switch on the lights or Play a Song in a regional language (Kannada, Telugu, Hindi, etc.).
4. Apply an SR approach that acknowledges the input received from the user.
5. Apply NLP to translate the input with NLT.
6. Examine the interpreted input for the local language; if not the local language, go to step number 3.
7. Robotization/Automation.
8. Acquire the reply to the input given from robotization across IoT.
9. Stock the data in a cloud platform or database (DB).

5.1. FLOWCHART OF THE APPROACH AND BLOCK DIAGRAM

The following figure (Block Diagram) depicts the abstraction of the proposed approach in the translation of speech recognized in the local language.



Block Diagram



Flowchart

The following diagram (Flowchart) depicts the sequence of steps performed in the proposed task.

6. IOT AUTOMATION

The popularity of cloud-based ideas and infrastructure has developed in recent years. Many different types of control applications in the realm of automation emphasize cloud computing. When it comes to automating tasks, an open Internet cloud paradigm is inadequate. Automation's physical and geographical locality means that automation clouds cannot be avoided^[8]. However, as mentioned^[9], local automated clouds should not rely on other resources to execute the specified automation functionalities. The Raspberry Pi, a computer that can run software like voice recognition and has Wi-Fi connectivity built-in^[10], is a popular platform for voice-enabled products. Therefore, the application software utilizes Google's speech and voice Application Programming Interfaces (APIs) to decode spoken input. This user's information is recognized and translated into natural language and transferred to a Raspberry Pi through the network, where it identifies and operates adjacent intelligent devices, and data is saved in a database.

7. CONCLUSIONS

Voice/Speech recognition with NLP and the help of the NLT, we translated the voice input and recognized with the robotization and has been automated quickly concerning the user's information. Text intake is not only in the form of a global language, English. However, it can be other Indian local languages like Kannada, Telugu, Hindi, etc.; subsequently, the voice is converted automatically and responded to in the same regional language.

REFERENCES

- [1] D. Machine Learning Paradigms for Speech Recognition: An Overview, Li Deng, Fellow, IEEE, and Xiao Li, Member, IEEE - IEEE Transactions on Audio, Speech, and Language Processing (Volume: 21, Issue: 5, May 2013)
- [2] Hidden Markov Models for Speech Recognition, B. H. Juang; L. R. Rabiner, *Technometrics*, 33:3, 251-272, DOI: 10.1080/00401706.1991.10484833
- [3] A hidden Markov model-based keyword recognition system Publisher: IEEE Cited by R.C. Rose; D.B. Paul - International Conference on Acoustics, Speech, and Signal Processing Date 3 -6 April 1990.
- [4] Introduction to Various Algorithms of Speech Recognition: Hidden Markov Model, Dynamic Time Warping and Artificial Neural Networks, Panini A. Trivedi V.V.P. Engineering College Rajkot, Gujarat, India-International Journal of Engineering Development and Research
- [5] Development of an automated English-to-local language translator using Natural Language Processing Iwara ARIKPO1, Iniobong, DICKSON2 - International Journal of Scientific & Engineering Research, Volume 9, Issue 7, July 2018, 378.
- [6] Speech Recognition for Light Control on Raspberry Pi Using Python Programming Paniti, Netinant College of Digital Innovation Technology, Rangsit University, Krairat, Meennapa, Rukhiran - ICSIM 2022: 2022 The 5th International Conference on Software Engineering and Information Management (ICSIM) January 2022 Pages 33–37.
- [7] Enabling IoT automation using local clouds - Jerker Delsing; Jens Eliasson; Jan van Deventer; Hasan Derhamy; Pal Varga -: 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT) Date of Conference: 12-14 Dec. 2016.
- [8] Automation using Artificial intelligence-based Natural Language processing P.Mohana; M. Muthuvinayagam; P.Umasankar; T. Muthumanickam, 2022, 6th International Conference on Computing Methodologies and Communication (ICCMC), Date of Conference: 29-31 March 2022.
- [9] Multi-modal pre-training for automated speech recognition - David M. Chan†1 Shalini Ghosh† Debmalya Chakrabarty† Björn Hoffmeister ICASSP 2022- 2022 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Date of Conference: 23-27 May 2022.
- [10] The Evaluation Process Automation of Phrase and Word Intelligibility Using Speech Recognition Systems - Salah, A., Karpov, A., Potapova, R. (eds) Speech, and Computer. SPECOM 2019. Lecture Notes in Computer Science (LNCS), vol 11658. Springer, Cham.
- [11] The State of the Art in Thai Language Processing Virach Sornlertlamvanich, Tanapong Potipiti, Chai WutiwWATCHAI and Pradit Mittrapiyanuruk National Electronics and Computer Technology Center (NECTEC), National Science and Technology Development Agency, Ministry of Science and Technology Environment. 22nd Floor Gypsum Metropolitan Tower 539/2 Sriyudhya Rd. Rajthevi Bangkok 10400 Thailand.
- [12] Speech Recognition using Neural Networks – Hardik Dudhrejia, Sanket Shah – International Journal of Engineering Research and Technology (IJRET) Volume 7, Issue 10, October 2018.
- [13] <https://www.speechmatics.com/resources/articles-and-news/speech-recognition-challenges-overcome>
- [14] <https://www.statista.com/statistics/1229799/india-smartphone-penetration-rate/#:~:text=Smartphone%20penetration%20rate%20in%20India%202010%2D2040&text=In%202020%2C%20the%20penetration%20rate,total%20population%20were%20using%20smartphone>