Review on Speech Recognition Using Natural Language Processing (NLP)

Shyam Narayan Singh
singhshyamnarayan9935@gmail.com

Khushi Srivastava
srivastavakhushi0609@gmail.com

Vinay Singh
vs67838@gmail.com

Mr. Arvind Kumar
omarvinda1812@gmail.com

Abstract

This era is the advanced technological era, speech recognition navigation systems play a major role in fulfilling the gap between humans and advanced machines. To overcome this difficulty in understanding the user's speech commands and natural language simulations. A system that recognizes and authenticates the voice of a user by extracting the different features of their voice samples is usually termed a Speech recognition system. Voice identification is carried out by compiling the human voice into machine-understanding language. The digitized audio samples then undergo a feature excerption process to extract Mel Frequency Cepstral Coefficients features. These coefficients are subjected to feature matching through Dynamic Time Warping to match with the patterns existing in the database for limited Hindi words. This paper focuses on a secure system that deploys speech recognition for a natural language (Hindi) by combining the digital and mathematical knowledge using MFCC and DTW to extract and match the features to improve the accuracy for better performance [1].

I. INTRODUCTION

With the enrichment of vast technologies, speech recognition has become very popular among Human and Machine Interfaces in recent years. This speech-based human-machine interaction creates a high value when the user interacts with the system. Speech Recognition is the emerging technology of security and authentication for the future. In the present day, confidential details encrypt by using text and image passwords are prone to attacks. In the case of the most commonly used text and image passwords, users are required to handle different passwords for emails, internet banking, and personal document. Hence users imply to choose passwords in such a way that they are easy to remember. But they are vulnerable in case of hackers. In the case of image passwords, they are vulnerable to shoulder surfing and other hacking techniques. Advances in speech technology have created a vast interest in the practical application of speech recognition that's why this system provides the users with the proper and efficient method of authentication system based on Speech recognition [2].

II. LITERATURE REVIEW

Humans are always comfortable communicating in their natural language. Communication is an integral part of human life, and also a symbol of identity and authorization. However, in the case of computers, human interaction, language accents and dialects differ for different sets of people [3].

Natural language processing (NLP) is a branch of computer science and more likely, the branch of artificial intelligence (AI) anxious with giving computers the ability to understand text and spoken words in a similar way that human beings can. NLP combines computational linguistics rule-based modeling of human language with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to 'understand' its full meaning, complete with the speaker or writer's intent and sentiment [9].
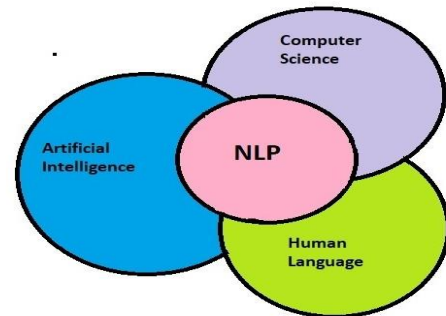

Fig 1: -Relation between AI, Computer science & Human language

NLP drives computer programs that translate text from one language to another language, respond to spoken commands, and summarize large volumes of text rapidly even in real-time. There's a good chance you've interacted with NLP in the form of voice-operated GPS systems, digital assistants, speech-to-text dictation software, customer service Chabot, and other consumer conveniences. But NLP also plays a growing role in enterprise solutions that help streamline business operations, increase employee productivity, and simplify mission-critical business processes.

NLP enables computers to understand natural language as humans do. Whether the language is spoken or written,

natural language processing uses artificial intelligence to take real-world input, process it, and make sense of it in a way a computer can understand. Just as humans have different sensors such as ears to hear and eyes to see computers have programs to read and microphones to collect audio. And just as humans have a brain to process that input, computers have a program to process their respective inputs. At some point in processing, the input is converted to code that the computer can understand [5,11].

There are two main phases to natural language processing: data preprocessing and algorithm development.
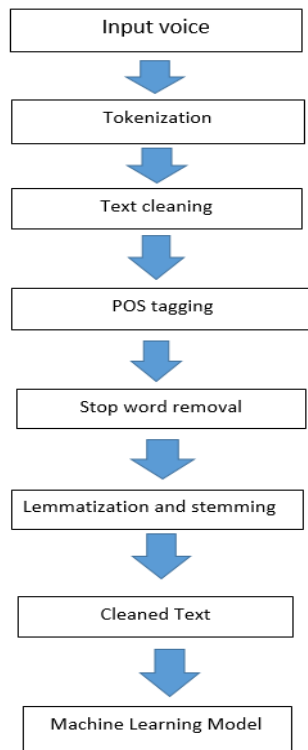

Fig 2: Steps of NLP

Data preprocessing involves preparing and "cleaning" text data for machines to be able to analyze it. preprocessing puts data in a workable form and highlights features in the text that an algorithm can work with.

There are several ways this can be done, including:
Tokenization. This is when text is divided into smaller units and work is done with these units.
Text cleaning. Text cleaning is the process of preparing raw text for NLP (Natural Language Processing) so that machines can understand human language
Stop word removal. This is when common words are removed from text so unique words that offer the most information about the text remain.
Lemmatization and stemming. This is when words are reduced to their root forms to process.

Part-of-speech tagging. This is when words are highlighted based on the part of speech they are such as nouns, verbs, and adjectives.

Once the data has been preprocessed, an algorithm is developed to process it. There are several natural language processing algorithms, but two main types are commonly used:

1. Rules-based system. This system follows pre-built designed linguistic rules. This approach was used early on in the development of natural language processing and is still used.

2. Machine learning-based system. Machine learning algorithms use statistical methods. They learn to perform tasks based on predefined training data and update their methods as more data is processed. Using a combination of machine learning, deep learning and neural networks, natural language processing algorithms have their own rules through repeated processing and learning

Cleaned data. Clean data are valid, accurate, complete, consistent, unique, and uniform

Machine Learning Model. Once you have trained the model, you can use it to reason over data that it hasn't seen before, and make predictions about those data. For example, let's say you want to build an application that can recognize a user's speech. You can train a model by providing human speech, and then you can use that model in an application that can recognize any user's speech.

History of NLP

The world of natural language processing starts in the 1940s, after World War II. At this time, people recognized the importance of translation from one language to another and hoped to create a machine that could do this sorting of translation automatically. However, the task was not as easy as people first imagined. By 1958, some researchers were identifying significant issues in the development of NLP. One of these researchers was Noam Chomsky, who found it troubling that models of language recognized sentences that were nonsense but grammatically correct as equally irrelevant as sentences that were nonsense and not grammatically correct. Chomsky found it problematic that the sentence "Colorless green ideas sleep furiously" was classified as improbable to the same extent that "Furiously sleep ideas green colorless"; any speaker of English can identify the former as grammatically correct and the latter as incorrect, and Chomsky felt the same should be expected of machine models.

Around the same time in history, from 1957-1970, researchers divided into two divisions concerning NLP: symbolic and stochastic. Symbolic, or rule-based, researchers focused on formal languages and generating syntax; this group consisted of many linguists and computer scientists who considered this branch the beginning of artificial intelligence research. Stochastic researchers were more interested in statistical and probabilistic methods of NLP, working on problems of optical character recognition and pattern recognition between texts.

After 1970, researchers subdivided even further, embracing new areas of NLP as more technology and knowledge became available. One new area was logic-based paradigms, languages that focused on encoding

rules and language in mathematical logic. This area of NLP research later contributed to the development of the programming language Prolog. Natural language understanding was another area of NLP that was particularly influenced by SHRDLU, Professor Terry Winograd's doctoral thesis. This program placed a computer in a world of blocks, enabling it to manipulate and answer questions about the blocks according to natural language instructions from the user. The amazing part of this system was its capability to learn and understand with amazing accuracy, something only currently possible in extremely limited domains (e.g., the block world). The following text was generated in a demonstration of SHDRLU:

Computer: OK. (does it)
Person: PUT THE LITTLEST TOWER ON TOP OF IT.
Computer: OK. (does it)
Person: DOES THE SHORTEST THING THE TALLEST TOWER'S SUPPORT SUPPORTS SUPPORT ANYTHING SKY BLUE?
Computer: YES, THE SKY BLUE TOWER.
Person: WHAT COLOR IS IT?
Computer: BY "IT", I ASSUME YOU MEAN THE SHORTEST THING THE TALLEST TOWER'S SUPPORT SUPPORTS.
Computer: SILVER

The computer can resolve relationships between objects and understand certain ambiguities. The fourth area of NLP that came into existence after 1970 is discourse modeling. This area examines interchanges between people and computers, working out such ideas as the need to change "you" in a speaker's question to "me" in the computer's answer.

From 1983 to 1993, researchers became more united in focusing on empiricism and probabilistic models. Researchers were able to test certain arguments by Chomsky and others from the 1950s and 60s, discovering that many arguments that were convincing in the text were not empirically correct. Thus, by 1993, probabilistic and statistical methods of handling natural language processing were the most common types of models. In the last decade, NLP has also become more focused on information extraction and generation due to the large amounts of information scattered across the Internet. Additionally, personal computers are now everywhere, and thus consumer-level applications of NLP are much more common and an impetus for further research [4].

III. Task of NLP

Human language is filled with ambiguities that make it incredibly difficult to write software that accurately determines the intended meaning of text or voice data. Homonyms, homophones, sarcasm, idioms, metaphors, grammar and usage exceptions, and variations in sentence structure—these just a few of the irregularities of human language that take humans years to learn, but that programmers must teach natural language-driven applications to recognize and understand accurately from the start if those applications are going to be useful.

Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it's ingesting. Some of these tasks include the following:

Speech recognition, also called speech-to-text, is the task of reliably converting voice data into text data. Speech recognition is required for any application that follows voice commands or answers spoken questions. What makes speech recognition especially challenging is the way people talk—quickly, slurring words together, with varying emphasis and intonation, in different accents, and often using incorrect grammar.

Part of speech tagging, also called grammatical tagging, is the process of determining the part of speech of a particular word or piece of text based on its use and context. Part of the speech identifies 'make' as a verb in 'I can make a paper plane,' and as a noun in 'What make of car do you own?'

Word sense disambiguation is the selection of the meaning of a word with multiple meanings through a process of semantic analysis that determine the word that makes the most sense in the given context. For example, word sense disambiguation helps distinguish the meaning of the verb 'make' in 'make the grade' (achieve) vs. 'make a bet' (place).

Named entity recognition, or NEM identifies words or phrases as useful entities. NEM identifies 'Kentucky' as a location or 'Fr0ed' as a man's name.

Techniques of NLP

In this section, the proposed recognition system is described. Voice recognition system upon which Google API acts as a platform. It is in-built pre-processing, segmentation, feature extraction, classification, and recognition.
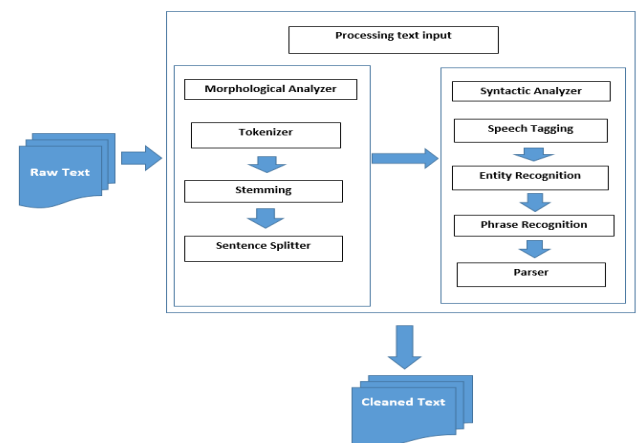

Fig 3: -Natural Language Processing Module

1-ASR

The goal of Automatic speech recognition (ASR) research is to address this problem computationally by building systems that map from an acoustic signal to a string of words. Automatic speech understanding(ASU) extends this goal to producing some sort of understanding of the sentence, rather than just words.

Fig 4: - Automatic Speech Recognition Module

It's a kind of technology that provides human beings to use their voice to speak with the machines like computer interfaces, which resemble normal human beings' conversations. Real-time streaming or pre-recorded audio input from an application's microphone can be streamed or sent from a particular pre-recorded audio file.

2-Auto detect language
This application detects the language spoken by the person by using google API (Application Program Interface) automatically.

3- Semantic Analysis-
Semantics involves the use of and meaning behind words. Natural language processing applies algorithms to understand the meaning and structure of sentences. Semantic techniques include:

Word sense disambiguation. This derives the meaning of a word based on context. Example: Consider the sentence, "The pig is in the pen." The word pen has different meanings. An algorithm using this method can understand that the use of the word pen here refers to a fenced-in area, not a writing implement.

Named entity recognition. This determines words that can be categorized into groups. Example: An algorithm using this method could analyze a news article and identify all mentions of a certain company or product. Using the semantics of the text, it would be able to differentiate between entities that are visually the same. For instance, in the sentence, "Daniel McDonald's son went to McDonald's and ordered a Happy Meal," the algorithm could recognize the two instances of "McDonald's" as two separate entities -- one a restaurant and one a person.

Natural language generation. This uses a database to determine the semantics behind words and generate new text. Example: An algorithm could automatically write a summary of findings from a business intelligence platform, mapping certain words and phrases to features of the data in the BI platform. Another example would be automatically generating news articles or tweets based on a certain body of text used for training.

Current approaches to natural language processing are based on deep learning, a type of AI that examines and uses patterns in data to improve a program's understanding. Deep learning models require massive amounts of labeled data for the natural language processing algorithm to train on and identify relevant correlations, and assembling this kind of big data set is one of the main hurdles to natural language processing.

Earlier approaches to natural language processing involved a more rules-based approach, where simpler machine learning algorithms were told what words and phrases to look for in text and given specific responses when those phrases appeared. But deep learning is a more flexible, intuitive approach in which algorithms learn to identify speakers' intent from many examples -- almost like how a child would learn human language.

4- syntax Analysis-
The syntax is the arrangement of words in a sentence to make grammatical sense. NLP uses syntax to assess meaning from a language based on grammatical rules. Syntax techniques include:

Parsing. This is the grammatical analysis of a sentence. Example: A natural language processing algorithm is fed the sentence, "The dog barked." Parsing involves breaking this sentence into parts of speech -- i.e., dog = noun, barked = verb. This is useful for more complex downstream processing tasks.

Word segmentation. This is the act of taking a string of text and deriving word forms from it. Example: A person scans a handwritten document into a computer. The algorithm would be able to analyze the page and recognize that the words are divided by white spaces.

Sentence breaking. This places sentence boundaries in large texts. Example: A natural language processing algorithm is fed the text, "The dog barked. I woke up." The algorithm can recognize the period that splits up the sentences using sentence breaking.

Morphological segmentation. This divides words into smaller parts called morphemes. Example: The word unconsciously would be broken into [[un[[conscious]able]]ly], where the algorithm recognizes "un," "conscious," "able" and "ly" as morphemes. This is especially useful in machine translation and speech recognition.

Stemming. This divides words with inflection in them to root forms. Example: In the sentence, "The lion roared," the algorithm would be able to recognize the root of the word "roared" is "roar." This would be useful if a user was analyzing a text for all instances of the word bark, as well as all of its conjugations. The algorithm can see that they are essentially the same word even though the letters are different [6,11].

IV. Application of NLP

1- Smart

e

A smart TV remote is one of the most popular examples of natural language processing. it follows the voice recognition concept through NLP. When we search for any content on our smart TV (like YouTube) we have to press the voice button of the TV remote and give voice input to it, after recognition of speech desired content is displayed on the smart TV as output [7,8,10].
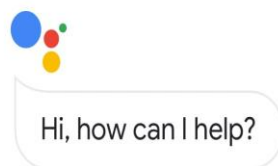
2- Google Assistant

Fig 5.b: - Google Assistant

Google Assistant works on the principle of NLP it recognizes the voice command and processed it to the desired output.
Users primarily interact with the Google Assistant through natural voice, though keyboard input is also supported. Assistant can answer questions, schedule events and alarms, adjust hardware settings on the user's device, show information from the user's Google account, play games, and more [7,8,10].

3-Grammarly
Grammarly Artificial intelligence (AI) system combines machine learning with a variety of natural language processing approaches. Human language has many levels at which it can be analyzed and processed: from characters and individual words through grammatical structures and sentences, even paragraphs or full texts [7,8,10].

4-Text Summarizer

Fig 5.c: - Text Summarizer

Text summarization is a very useful and important part of Natural Language Processing (NLP). First, let us talk about what text summarization is. Suppose we have too many lines of text data in any form, such as from articles or magazines or on social media. We have time scarcity so we want only a nutshell report of that text. We can summarize our text in a few lines by removing unimportant text and converting the same text into a smaller semantic text form [7,8,10].

V. Conclusion
In this section, we have studied NLP tasks, techniques, history, and literature review which give very vast knowledge about voice recognition using NLP there are some figures which help us to understand it more.
Natural Language Processing is the process of teaching machines to understand and interpret conversational inputs from humans. NLP based on Machine Learning can be used to establish communication channels between humans and machines. Natural language processing is one of the categories of computer science and AI that focuses mainly on the interaction between computers and humans. The very first NLP was designed in 1950. Some real-life application of Natural language processing includes Apple's, Siri and Microsoft's Cortana.
Extracting the main points from a text is a familiar problem in the field of machine learning (ML) and natural language processing (NLP) research, but our unique use case and constraints led us to apply a different approach to the problem.

VI. References
1. Pooja Withanage, Tharaka Liyanage, Naditha Deeyakaduwe, Eshan Dias, Samantha Thelijjagoda, "Using Voice-based Road Navigation System Natural Language Processing (NLP)". 978-1-5386-9418-3/18/$31.00 ©2018 IEEE
2. Erik Cambria, Bebo White, "Jumping NLP Curves: A Review of Natural Language Processing Research", IEEE Computational intelligence magazine, 1556-603x/14/$31.00©2014IEEE, Date of publication:11 April 2014
3. W. Bright, "Natural language processing: an overview". Oxford University Press,

1992, Vol. 3, 53-59.

4. Karen Sparck Jones, "Natural Language Processing: A Historical Review", Springer Science+Business Media Dordrecht 1994

5. Dr. Kavitha. R , Nachammai., Ranjani., Shifali, "Speech Based Voice Recognition System for Natural Language Processing", Kavitha. R et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5301-5305, ISSN:0975-9646

6. REVIEW ON NATURAL LANGUAGE PROCESSING Prof. Alpa Reshamwala, Prof. Dhirendra Mishra, Prajakta Pawar,

7. "REVIEW ON NATURAL LANGUAGE PROCESSING", IRACST – Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.3, No.1, February 2013

8. Mrs. Paul Jasmin Rani, Jason Bakthakumar, Praveen Kumaar, Praveen Kumaar and Santhosh Kumar, "VOICE CONTROLLED HOME AUTOMATION SYSTEM USING NATURAL LANGUAGE PROCESSING (NLP) AND INTERNET OF THINGS (IoT)", 2017Third International Conference on Science Technology Engineering & Management (ICONSTEM), 978-1-5090-4855-7/17/$31.00 ©2017 IEEE.

9. George Alexakis, Spyros Panagiotakis, Alexander Fragkakis, Evangelos Markakis and Kostas Vassilakis "Control of Smart Home Operations Using Natural Language Processing, Voice Recognition and IoT Technologies in a Multi-Tier Architecture". Received: 15 April 2019; Accepted: 14 June 2019; Published: 1 July 2019.,

10. Stephen W. Kercel and Raymond W. Tucker, "SPEAKER RECOGNITION THROUGH NLP AND CWT MODELING" 15th Annual Security Technology Symposium Security-Related Research and Methodology Session Norfolk, Virginia June 16, 1999.

11. Dwijen Rudrapal Smita Das, S. Debbarma N. Kar, N. Debbarma "Voice Recognition and Authentication as a Proficient Biometric Tool and its Application in Online Exam for P.H People". International Journal of Computer Applications (0975 – 8887) Volume 39– No.12, February 2012.

12. J. Pradeep, K. Vijayakumar, and M. Harikrishnan "Voice Recognition Using Natural Language Processing" E-mail: pradeepjesmvec.ac.in; msg2world@gmail.com: harikrishnan@smvec.ac.in

13. Daniel Jurafsky, James H..Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition", Second edition, Pearson, 2e Paperback