# SPEECH TO TEXT USING DEEP LEARNING

[1] L Nalini Joseph, [2] Balu Madhu , [3] L Vamsi Krishna Reddy , [4] M Raj Kumar, [5] S Uday Kiran,

[1]Professor From Bharath Institute Of Higher Education And Research, Chennai- 600073

[2,3,4,5] Students of Computer Science and Engineering ,Bharath Institute Of Higher Education And Research, Chennai-600073,

***ABSTRACT: Speech recognition is one of the fastest growing engineering technologies. It has many applications in various fields, and offers many potential benefits. Many people could not communicate due to the language barrier. Our goal is to reduce this barrier with our program designed and developed to access the system in specific cases, providing vital assistance in enabling people to share information by operating the system using voice input Could. This project takes that factor into account and strives to ensure that our program recognizes speech and converts the input audio to text; This enables the user to perform file operations such as save, open, or opt out of voice-only input. We design a system that recognizes human voices and audio clips and translates between English and English. The output is in text format and we provide options to convert the audio from one language to another. Next, we hope to add a function that provides dictionary meanings for English words. Neural machine translation is the primary method used to perform machine translation in the industry. This work on speech recognition begins with an introduction to the technology and its applications in various fields. Part of the report is based on software improvements in speech recognition.***

***Keywords: Speech Recognition, Communicate, Input, Text, Language, Neural Machine Translation***

## *INTRODUCTION*

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. Many speech recognition applications, such as voice dialing, simple data entry and speech-to-text are in existence today. Automatic speech recognition systems involve numerous separate components drawn from many different disciplines such as statistical pattern recognition, communication theory, signal processing, combinatorial mathematics, and linguistics. Speech recognition is an alternative to traditional methods of interacting with a computer, Like text input through keyboard. Attempts to develop an effective system that could replace or reduce the reliability of standard keyboard input, automatic speech recognition (ASR) systems, were first attempted in the 1950s. These early speech recognition systems attempted to use a set of grammatical and syntactic rules for speech recognition. If the spoken words conform to a certain rule, the computer can recognize the words. However, human language has many exceptions to its rules. The way words and phrases are spoken can be substantially changed by accent, dialects, and customs. Therefore, we use algorithms to obtain ASR. In modern civilized societies, the most common method of communication between humans is speech. Various ideas formed in the mind of the speaker are expressed through speech in the form of words, phrases and sentences by applying certain correct grammatical rules. Speech is the primary medium of communication between humans, and speech is the most natural and efficient form of communication between humans. By classifying speech by sounds, sounds, and silences (VAS/S), we can consider the basic acoustic segmentation required for speech. By following individual sounds called phonemes, this technique closely resembles the sounds of each letter of the alphabet that make up the structure of human speech. The main objective of speech recognition is to generate a set of words from a sound signal received from a microphone or telephone. Computers must be used to extract and determine the linguistic information communicated by speech waves.

## *PROBLEM STATEMENT:*

The main objective of our project is to encourage the use of our native language and to help illiterate people for the easy typing of the text. The concept involves the recognition of voice through a microphone. The background noise is reduced by using a noise filter cap over the mic. The words are extracted from the input voice by using some artificial intelligence and feature extraction. By using NLTK we convert the word and these words in the uttered speech are identified using word tokenizer. The extracted words are then compared with the pre-trained data set by using data analysis.

## *LITERATURE REVIEW*

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system. The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

***Real time text-to speech conversion system for Spanish***

Research aiming to develop a text-to-speech converter (TSC) for Spanish, that accepts a continuous source of alphanumeric characters (up to 250 words per minute) and produces good quality, natural Spanish output, is described. Four sets of problems are considered in this work: the hard-ware structure adopted for real time operation; the complex control software needed to handle the orthographic input and linguistic programs; the linguistic processing rules, and the parametrization of the Spanish language matched to a TSC. Emphasis is made on the problems of adapting a general hardware structure to a specific language.

***Hindi-English Speech-to-Speech Translation System for Travel Expressions***

Speech-to-speech translation system enables in translation of speech signals in a source language A to target language B. A good speech-to-speech translation (S2ST) system can be characterized by its ability to keep intact the fluency and meaning of the original speech input. An S2ST system to enable translation between Hindi and English is the main idea of the proposed work. A preliminary dataset concentrating on basic travel expressions in both the languages considered is used for this work. In order to develop a successful S2ST system three subsystems are required namely, automatic speech recognition (ASR) system, machine translation (MT) system and text-to-speech synthesis (TTS) system. Hidden Markov models based ASR system is developed for both the languages and their performances are analyzed based on the word error rate (WER). The MT subsystem makes use of the statistical machine translation (SMT) approach for the purpose of translating the text between the two languages involved. The SMT makes use of IBM alignment models and language models to enable proper translation. The performance of MT is analyzed based on translated edit rate (TER) and analysis of the translation table. HMM-based speech synthesis system (HTS) is used to synthesize the translated text. Performance of the synthesizer is analyzed based on mean opinion score (MOS) from a group of listeners.

***Acoustic modeling problem for automatic speech recognition system***

In automatic speech recognition (ASR) systems, the speech signal is captured and parameterized at front end and evaluated at back end using the statistical framework of hidden Markov model (HMM). The performance of these systems depend critically on both the type of models used and the methods adopted for signal analysis. Researchers have proposed a variety of modifications and extensions for HMM based acoustic models to overcome their limitations. In this review, we summarize most of the research work related to HMM-ASR which has been carried out during the last three decades. We present all these approaches under three categories, namely conventional methods, refinements and advancements of HMM. The review is presented in two parts (papers): (i) An overview of conventional methods for acoustic phonetic modeling, (ii) Refinements and advancements of acoustic models. Part I explores the architecture and working of the standard HMM with its limitations. It also covers different modeling units, language models and decoders. Part II presents a review on the advances and refinements of the conventional HMM techniques along with the current challenges and performance issues related to ASR.

***Speech recognition for English to Indonesian translator using Hidden Markov Model***

Communication is needed for the process of delivering information. Language is used as a communication for interacting between humans. But not everyone has the same language in communicating, there are many languages from different countries in the world. This paper describes a system that can translate a language. The translated language is English to Indonesian based on speech recognition with feature extraction using Mel Frequency Cepstral Coefficients (MFCC) and Hidden Markov Model (HMM) classification method.

***Automatic Speech-Speech translation form of English language and translate into Tamil language***

Nowadays in most of the places people face problems while speaking with other people, who does not know their languages or other languages, in existing system and technique they record the interaction and translate into another language using manual transaction, in order to avoid the difficulties, In this paper we are going to automatically recognize the speech in the form of English language and translate into Tamil language. In this device it consists of three parts namely speech recognition device, English to Tamil machine translation and Tamil speech generation, it first recognize the speech (in English) using speech reorganization device and it displayed it on the screen in English (text) then translate into Tamil language (text) and displayed it, after that it convert into Tamil speech and it should be heard at the other end of the device.

***Multilingual speech-to-speech translation system for mobile consumer devices***

Along with the advancement of speech recognition technology and machine translation technology in addition to the fast distribution of mobile devices, speech-to-speech translation technology no longer remains as a subject of research as it has become popularized throughout many users. In order to develop a speech-to-speech translation system that can be widely used by many users, however, the system needs to reflect various characteristics of utterances by the users who are actually to use the speech-to-speech translation system other than improving the basic functions under the experimental environment. This study has established a massive language and speech database closest to the environment where speech-to- speech translation device actually is being used after mobilizing plenty of people based on the survey on users' demands. Through this study, it was made possible to secure excellent basic performance under the environment similar to speech-to-speech translation environment, rather than just under the experimental environment. Moreover, with the speech-to-speech translation UI, a user-friendly UI has been designed; and at the same time, errors were reduced during the process of translation as many measures to enhance user satisfaction were employed. After implementing the actual services, the massive database collected through the service was additionally applied to the system following a filtering process in order to procure the best-possible robustness toward both the details and the environment of the users' utterances. By applying these measures, this study is to unveil the procedures where multi-language speech-to-speech translation system has been successfully developed for mobile devices.

### *SOPC-Based Speech-to-Text Conversion*

For the past several decades, designers have processed speech for a wide variety of applications ranging from mobile communications to automatic reading machines. Speech recognition reduces the overhead caused by alternate communication methods. Speech has not been used much in the field of electronics and computers due to the complexity and variety of speech signals and sounds. However, with modern processes, algorithms, and methods we can process speech signals easily and recognize the text.

### *Text-to-speech algorithms based on FFT synthesis*

Present FFT synthesis algorithms for a French text-to-speech system based on diaphone concatenation. FFT synthesis techniques are capable of producing high quality prosodic adjustments of natural speech. Several different approaches are formulated toreduce the distortions due to diaphone concatenation.

### *Explore Speech-to-Text(STT) conversion using SAPI for Bangla language*

Although achieved performance is promising for STT related studies, they identified several elements to recover the performance and might give better accuracy and assure that the theme of this study will also be helpful for other languages for Speech-to-Text conversion and similar tasks.

## *OBJECTIVE:*

The main objective of our project is to encourage the use of our native language and to help illiterate people for the easy typing of the text. The concept involves the recognition of voice through a microphone. The background noise is reduced by using a noise filter cap over the mic. The words are extracted from the input voice by using some artificial intelligence and feature extraction. By using NLTK we convert the word and these words in the uttered speech are identified using word tokenizer. The extracted words are then compared with the pre-trained data set by using data analysis.

## *EXISTING SYSTEM*

The Java Speech API defines a standard, cross platform software interface to state-of-the-art speech technology. Two core speech technologies are supported through the Java Speech API: speech recognition and speech synthesis. Speech recognition provides computers with the ability to listen to spoken language and to determine what has been said. In other words, it processes audio input containing speech by converting it to text. The Java Speech API was developed through an open development process. With the active involvement of leading speech technology companies, with input from application developers and with months of public review and comment, the specification has achieved a high degree of technical excellence. As a specification for a rapidly evolving technology, Sun will support and enhance the Java Speech API to maintain its leading capabilities.The Java Speech API is an extension to the Java platform. Extensions are packages of classes written in the Java programming language (and any associated native code) that application developers can use to extend the functionality of the core part of the Java platform.But the main disadvantage is it recognize only the some reserved words only

### *Disadvantages:*

1. some solutions don’t easily integrate with existing HER.
2. some operating systems (namely Macs) and devices have less support for speech recognition
3. some solutions don’t have many customization options

## *PROPOSED SYSTEM*

Speech recognition aids in translating the spoken language into text. We have come up with a Speech Recognition model that converts the speech data given by the user as an input into the text format in his desired language. This model is developed by adding Multilingual features to the existent Google Speech Recognition model based on some of the natural language processing principles. The goal of this research is to build a speech recognition model that even facilitates an illiterate person to easily communicate with the computer system in his regional language.

### *Advantages:*

1. Solves Inefficiencies and Reduces Wasted Time

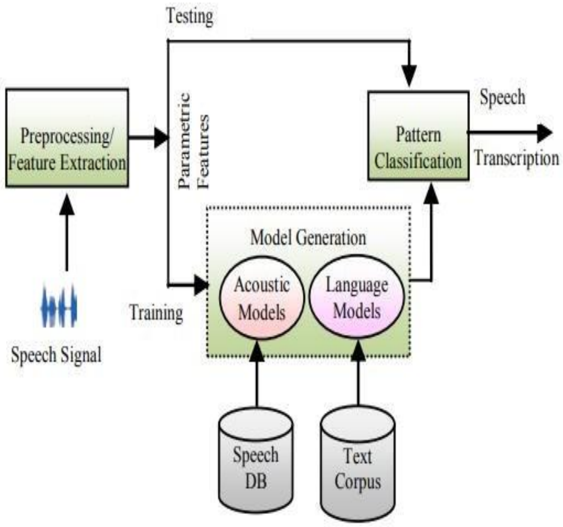
2. Clinics and Hospitals Can Save Money

3. Clinician Satisfaction

## *GOALS:*

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.

7. Integrate best practices.



*SYSTEM* ARCHITECTURE

a. Easy to measure extracted speech features.

b. It should not be susceptible to mimicry.

c. It should show little fluctuation from one speaking environment to another.

d. It should be stable over time.

E .It should occur frequently and naturally in speech.

***Speech To Text Module***

A real time speech to text conversion system converts the spoken words into text form exactly in the similar way that the user pronounces. We created a real time speech recognition system that was tested in real time noiseous environment. The purpose of this project was to introduce a new speech recognition system that is computationally simple and more robust to noise than the HMM based speech recognition system.

## *SYSTEM REQUIREMENTS*

### *HARDWARE REQUIREMENTS*

- System: Pentium Dual Core.
- Hard Disk: 120 GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram: 4 GB.

### *SOFTWARE REQUIREMENTS*

- Operating system: Windows 7/10.
- Coding Language :Python

## *MODULES*

1. Speech Analysis Module
2. Feature Extraction Module
3. Speech To Text Module

### *MODULE DESCRIPTIONS*

### *Speech Analysis Module*

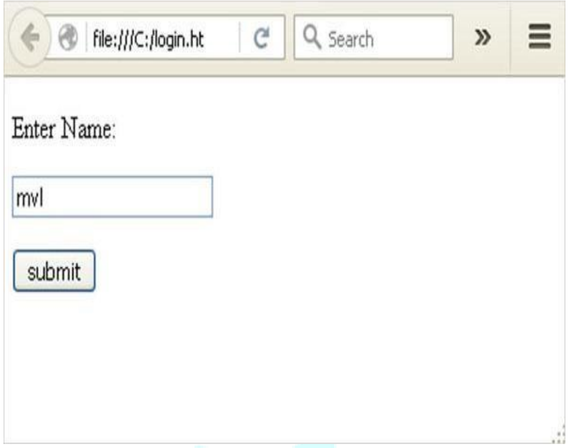
In Speech analysis technique Speech data contains different types of information that shows a speaker identity. This includes speaker specific information due to vocal tract, excitation source and behavior feature. The physical structure and dimension of vocal tract as well as excitation source are unique for each speaker. This uniqueness is embedded in the speech signal during speech production and can be used for speaker used for speaker recognition.

### *Feature Extraction Module*

Feature Extraction is the most important part of speech recognition since it plays an important role to separate one speech from other. Because every speech has different individual characteristics embedded in utterances. These characteristics can be extracted from a wide range of feature extraction techniques proposed and successfully exploited for speech recognition task. But extracted feature should meet some criteria while dealing with the speech signal such as:

welcome mvl

## *RESULT and DISCUSSION*



Enter Name:

mvl

submit

Form data is POSTed to the URL in action clause of form tag.

**http://localhost/login** is mapped to the **login()** function. Since the server has received data by **POST** method, value of 'nm' parameter obtained from the form data is obtained by −

```
user = request.form['nm']
```

It is passed to **'/success'** URL as variable part. The browser displays a **welcome** message in the window.

Change the method parameter to **'GET'** in **login.html** and open it again in the browser. The data received on server is by the **GET** method. The value of 'nm' parameter is now obtained by −

```
User = request.args.get('nm')
```

Here, **args** is dictionary object containing a list of pairs of form parameter and its corresponding value. The value corresponding to 'nm' parameter is passed on to '/success' URL as before.

### *GOALS:*

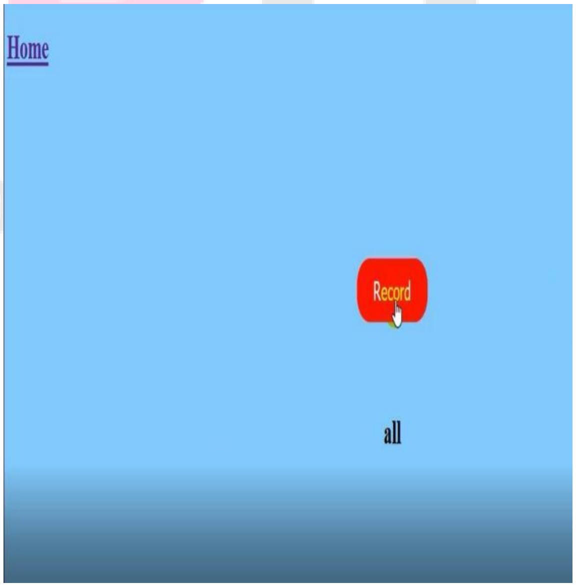
The Primary goals in the design of the UML are as follows:

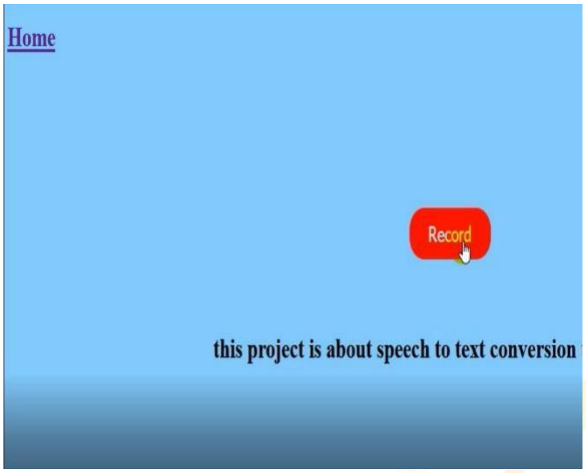
1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
8. Integrate best practices.

### *CONCLUSION*

By implementing this model, we learned how speech recognition packages can be used to build speech translation models. The more we use these types of packages, the more flexibility we get in code display and output. This model can be used for any purpose of speech to text translation. This model has many advantages, one of them is that you can live in unknown places where you do not know the speaking language, but with the help of this model you can translate that regional speech into text and in areas like Can also use it. Telecommunication. and multimedia. Additionally, this model is also useful for providing effective communication between man and machine

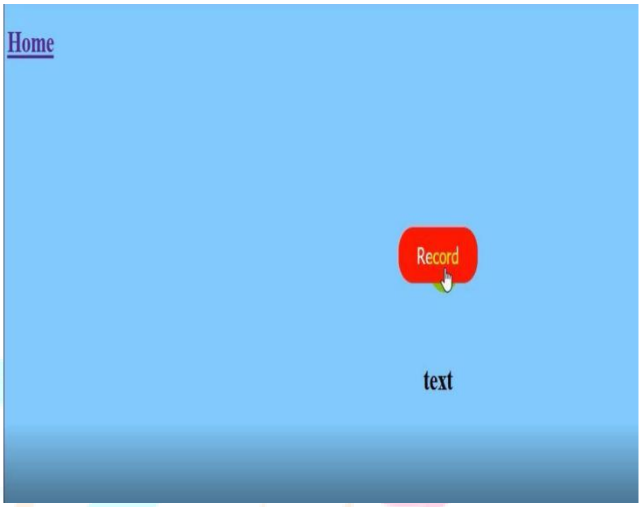
### *SCREENSHOTS*





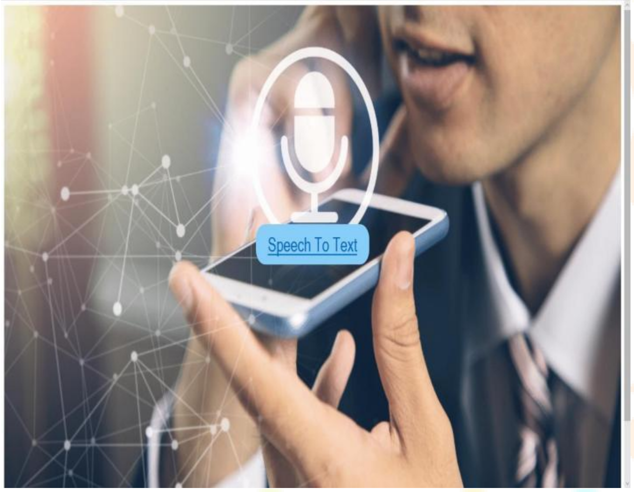
this project is about speech to text conversion

*Speech to Text Page*



text

*Home Page*



## *REFERENCES*

[1] Mrinalini Ket al: Hindi-English Speech-to-Speech Translation System for Travel Expressions, 2015 International Conference on Computation of Power, Energy, Information And Communication.

[2] Development and Application of Multilingual Speech Translation Satoshi Nakamura', Spoken Language Communication Research Group Project, National Institute of Information and Communications Technology, Japan.

[3] Speech-to-Speech Translation: A Review, Mahak Dureja Department of CSE The NorthCap University, Gurgaon Sumanlata Gautam Department of CSE The NorthCap University, Gurgaon. International Journal of Computer Applications (0975 – 8887) Volume 129 – No.13, November2015.

[4] Sequence-to-Sequence Models for Emphasis Speech Translation. Quoc Truong Do,Skriani Sakti; Sakriani Sakti; Satoshi Nakamura, 2018 IEEE/ACM.

[5] Olabe, J. C.; Santos, A.; Martinez, R.; Munoz, E.; Martinez, M.; Quilis, A.; Bernstein, J., “Real time text-to speech conversion system for spanish," Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84. , vol.9, no., pp.85,87, Mar 1984.

[6] Kavaler, R. et al., “A Dynamic Time Warp Integrated Circuit for a 1000-Word Recognition System”, IEEE Journal of Solid-State Circuits, vol SC-22, NO 1, February 1987, pp 3-14.

[7] Aggarwal, R. K. and Dave, M., “Acoustic modeling problem for automatic speech recognition system: advances and refinements (Part II)”, International Journal of Speech Technology (2011) 14:309–320.

[8] Ostendorf, M., Digalakis, V., & Kimball, O. A. (1996). “From HMM’s to segment models: a unified view of stochastic modeling for speech recognition”. IEEE Transactions on Speech and Audio Processing, 4(5), 360– 378.

[9] Yasuhisa Fujii, Y., Yamamoto, K., Nakagawa, S., “AUTOMATIC SPEECH RECOGNITION USING HIDDEN CONDITIONAL NEURAL FIELDS”, ICASSP 2011: P-5036-5039.

[10] Mohamed, A. R., Dahl, G. E., and Hinton, G., “Acoustic Modelling using Deep Belief Networks”, submitted to IEEE TRANS. On audio, speech, and language processing, 2010.