PROBLEM STATEMENT

Design an intelligent algorithm leveraging **Natural Language Processing**, **Speech Recognition**, **Big-Data/Machine-Learning** techniques that can transliterate and translate spoken words from a recording (Audio) and generate a transcript in multiple **Indian Languages** (Initially English and Hindi and Later Kannada/Telugu).

CHAPTER 1

INTRODUCTION

1.1 PROJECT OBJECTIVE

- To understand the speech recognition and its fundamentals
- It's working and application in different areas
- It's implementation as desktop application / website
- This software can be mainly used for
  - Speech Recognition
  - Speech generation
  - Text generation
  - Speech to text conversion
  - Transliterate obtained text to local regional language

1.2 ABSTRACT

Voice is the basic, common and efficient form of communication method for people to interact with each other. This project develops a service in which human voice is recognized by machine and it is converted to text of various Indian Languages. Our project mainly focuses on the voice recognizing and converting acoustic signal captured using microphone to a set of words. The recorded data can be used for documentation purposes.

Speech recognition technology is one from the fast growing engineering technologies. It has number of applications in different areas and provides a potential benefits. Speech Recognition is the ability of a computer to recognize general, naturally flowing utterances from a wide variety of users. It recognizes the caller's answers to move along the flow of the call. Nearly 20% of people in the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. The speech recognition systems in those particular cases provide a significant help to them, so that they can share information with people by operating computer through voice input. While there is still much room for improvement, current speech recognition systems have remarkable performance. We are only humans, but as we develop this technology and build remarkable changes we attain certain achievements. Rather than asking what is still deficient, we ask instead what should be done to make it efficient.

Speech-to-text allows computer to translate voice request and dictation into text. Voice recognition system: speech-to-text is the process of converting acoustic signal which is

captured using a microphone to set of words.

1.3 PROJECT SCOPE

1. It helps in applying machine learning concepts in Real Time Scenarios.
2. It helps us in gaining proficiency in designing solutions to problems.
3. It helps us in understating of the domain and learn concepts of Hidden Markov Models
4. It helps us in understanding the programming tools used in modern computer age.
5. This project can be used to transcribe pre-recorded audio to various Indian Languages.

CHAPTER 2

LITERATURE SURVEY

2.1 AN OVERVIEW OF SPEECH RECOGNITION

Speech Recognition is a technology that able a computer to capture the words spoken by a human with the help of microphone. These words are later recognized by speech recognizer, and in the end, the system captures recognized words. Further the words are transliterated to different local languages and output of transliterated words is shown on the screen. The process of speech recognition consists of different steps which will be discussed in the following sections one by one.

An ideal situation in the process of speech recognition is that, a speech recognition engine recognizes all words uttered by a human but, practically the performance of a speech recognition engine depends on number of factors. Vocabularies, multiple users and noisy environment are the major factors that are counted in as the depending factors for a speech recognition engine.

2.2 History

The concept of speech recognition started somewhere in 1940s, practically the first speech recognition program was appeared in 1952 at the bell labs, that was about recognition of a digit in a noise free environment.

After the five decades of research, the speech recognition technology has finally entered marketplace, benefiting the users in variety of ways. The challenge of designing a machine that truly functions like an intelligent human is still a major one going forward.

2.3 Types of Speech Recognition

Speech recognition systems can be divided into the number of classes based on their ability to recognize that words and list of words they have. A few classes of speech recognition are classified as follows:

2.3.1 Isolated Speech

Isolated words usually involve a pause between two utterances; it doesn't mean that it only accepts a single word but instead it requires one utterance at a time.

2.3.2 Connected Speech

Connected words or connected speech is similar to isolated speech but allow separate utterances with minimal pause between them.

2.3.3 Continuous Speech

Continuous speech allow the user to speak almost naturally, it is also called the computer dictation

2.3.4 Spontaneous Speech

At a basic level, it can be thought of as speech that is natural sounding and not rehearsed. An ASR system with spontaneous speech ability should be able to handle a variety of natural speech features such as words being run together, "ums" and "ahs", and even slight stutters.

2.4 Speech Recognition Process

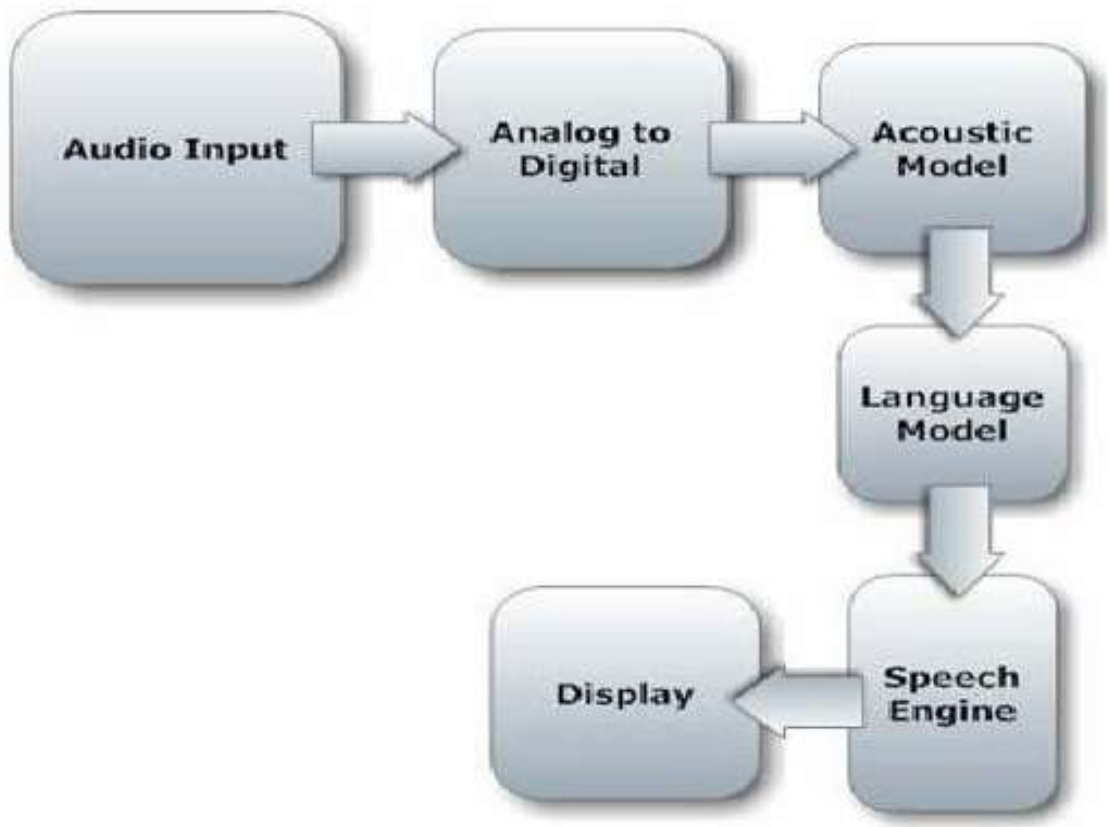

Fig 2.1 Speech Recognition Overview

2.5 Components of Speech Recognition System

2.5.1 Voice Input

With the help of microphone audio is input to the system, the pc sound card produces the equivalent digital representation of received audio

2.5.2 Digitization

The process of converting the analog signal into a digital form is known as digitization, it involves the sampling and quantization processes. Sampling is converting a continuous signal into discrete signal, while the process of approximating a continuous range of values is known as quantization

2.5.3 Acoustic Model

An acoustic model is created by taking audio recordings of speech, and their text transcriptions, and using software to create statistical representations of the sounds that make up each word. It is used by a speech recognition engine to recognize speech. The software acoustic model breaks the words into the phonemes

2.5.4 Language Model

Language modeling is used in many natural language processing applications such as speech recognition tries to capture the properties of a language and to predict the next word in the speech sequence. The software language model compares the phonemes to words in its built in dictionary.

2.5.5 Speech Engine

The job of speech recognition engine is to convert the input audio into text; to accomplish this it uses all sorts of data, software algorithms and statistics. Its first operation is digitization as discussed earlier, that is to convert it into a suitable format for further processing. Once audio signal is in proper format it then searches the best match for it. It

does this by considering the words it knows, once the signal is recognized it returns its corresponding text string.

2.6 Applications

2.6.1 Medical perspective

People with disabilities can benefit from speech recognition programs. Speech recognition is especially useful for people who have difficulty using their hands, in such cases speech recognition programs are much beneficial and they can use for operating computers. Speech recognition is used in deaf telephony, such as voicemail to text.

2.6.2 Military perspective

Speech recognition programs are important from military perspective; in Air Force speech recognition has definite potential for reducing pilot workload. Beside the Air force such programs can also be trained to be used in helicopters, battle management and other applications.

2.6.3 Educational perspective

Individuals with learning disabilities who have problems with thought-to-paper communication (essentially they think of an idea but it is processed incorrectly causing it to end up differently on paper) can benefit from the software. Some other application areas of speech recognition technology are described as below

i) Command Control
ii) Telephony
iii) Medical Disabilities

2.7 Future of speech recognition

- Accuracy will become better and better.
- Dictation speech recognition will gradually become accepted.
- Greater use will be made of “intelligent systems” which will attempt to guess what the speaker intended to say, rather than what was actually said, as people often misspeak and make unintentional mistakes.
- Microphone and sound systems will be designed to adapt more quickly to changing background noise levels, different environments, with better recognition of extraneous material to be discarded.

2.8 Speech Recognizing software

2.8.1 CMU Sphinix

Sphinx originally started at CMU and has recently been released as open source. This is a fairly large program that includes a lot of tools and information. It is still "in development", but includes trainers, recognizers, acoustic models, language models, and some limited documentation. This software is primarily for developers.

Homepage: http://www.speech.cs.cmu.edu/sphinx/Sphinx.html

Source: http://download.sourceforge.net/cmusphinx/sphinx2−0.1a.tar.gz

2.9 Speech Translation

Speech is an exceptionally attractive modality for human computer interaction: it is “hands free”; it requires only modest hardware for acquisition (a high quality microphone or microphones); and it arrives at a very modest bit rate Recognizing human speech, especially continuous (connected) speech, without burdensome training spacer independent), for a vocabulary of sufficient complexity (60,000 words) is very hard However with modern processes, flow diagram, algorithms, and methods we can process speech signals easily and recognize the text which is talking by the talker.

In this system, we are going to develop an online speech to text engine. The system acquires speech at run time through a microphone and processes the sampled speech to identify the

uttered text. The recognized text can be stored in a file. It can supplement other larger systems, giving users a different choice for data entry

2.10 Vocabulary

The vocabulary size of speech recognition system affects the processing requirements, accuracy and complexity of the system. In voice recognition system: speech-to-text the types of vocabularies can be classified as follows:

1) Small vocabulary: single letter.

2) Medium vocabulary: two or three letter words.

3) Large vocabulary: more letter words.

2.11 Grammars

A grammar describes a very simple type of the language for command and control. They are usually written by hand or generated automatically within the code. Grammars usually do not have probabilities for word sequences, but some elements might be weighed. They can be created with the Java Speech Grammar Format (JSGF) and usually have a file extension like *.gram* or *.jsgf*.

Grammars allow you to specify possible inputs very precisely, for example, that a certain word might be repeated only two or three times. However, this strictness might be harmful if your user accidentally skips the words which the grammar requires. In that case the whole recognition will fail. For that reason it is better to make grammars more flexible. Instead of phrases, just list the bag of words allowing arbitrary order. Avoid very complex grammars with many rules and cases. It just slows down the recognizer and you can use simple rules instead. In the past, grammars required a lot of effort to tune them, to assign variants properly and so on. The big VXML consulting industry was about that.

CHAPTER 3

DETAILED DESIGN

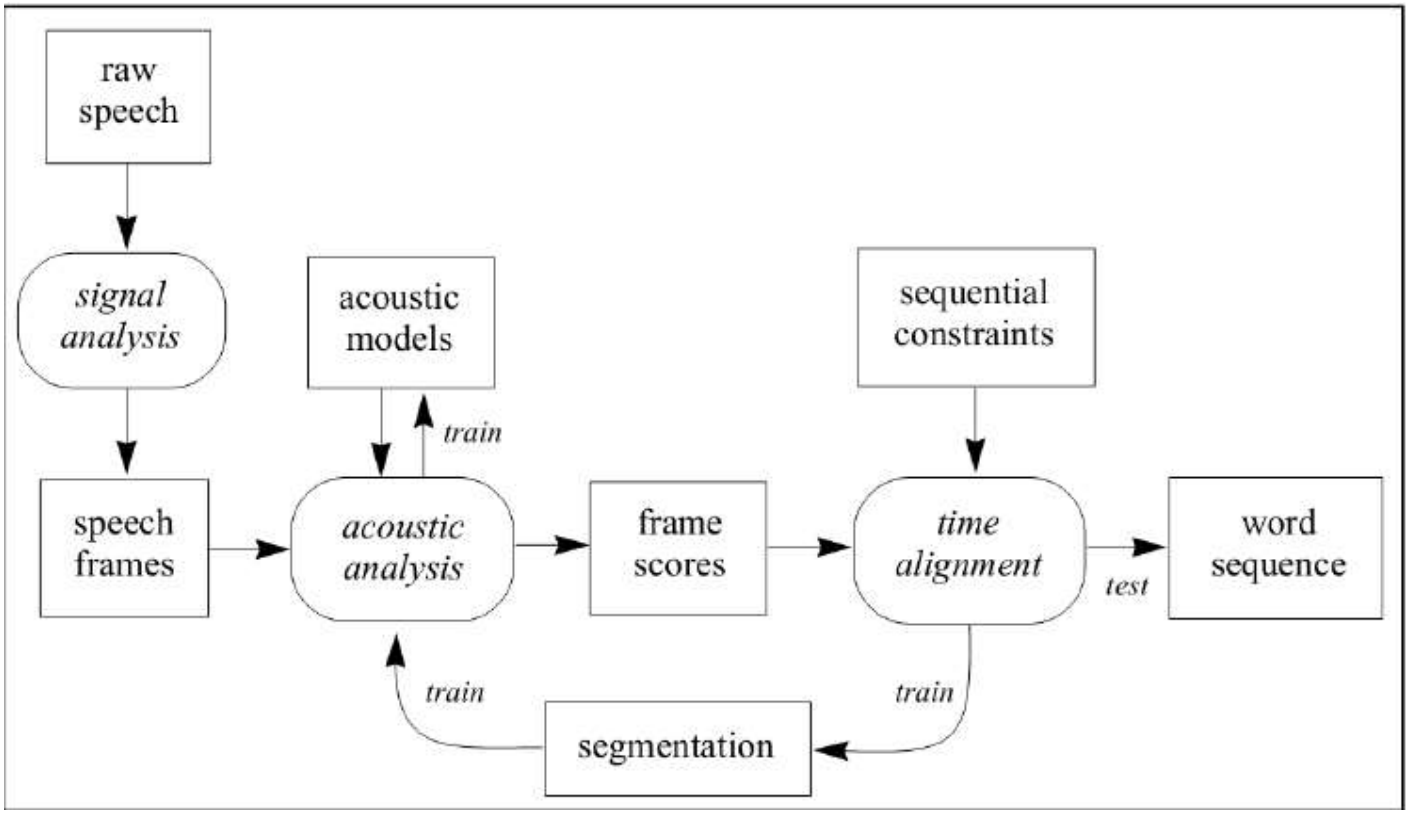

Fig 3.1 Speech Recognition System

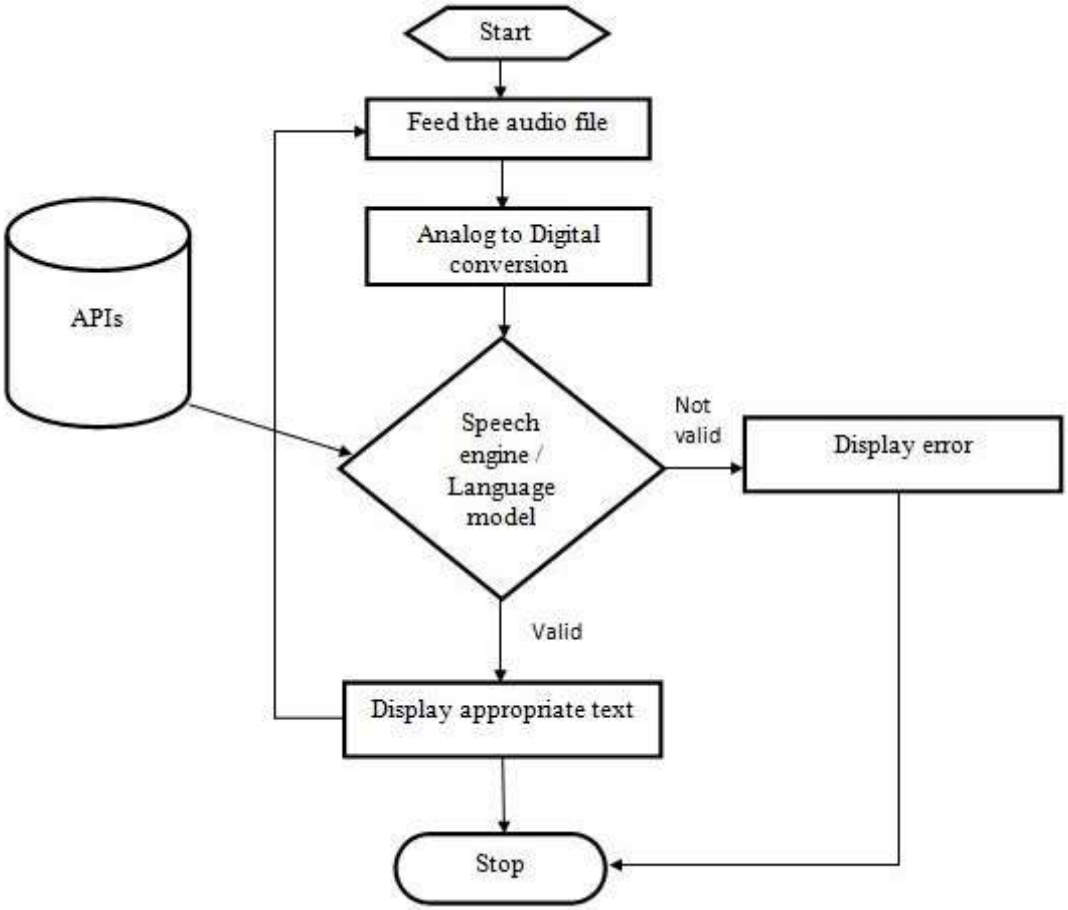

Fig 3.2 Block Diagram of Prototype

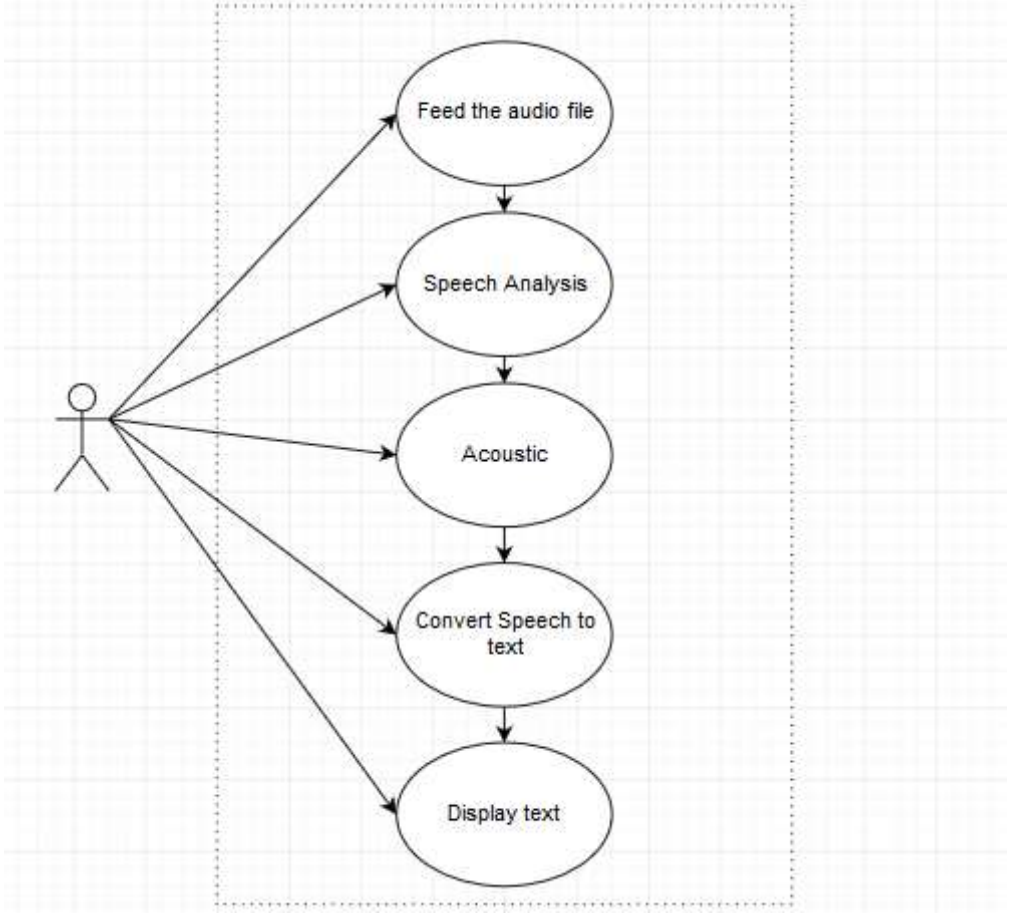

Fig 3.3 Use Case Diagram

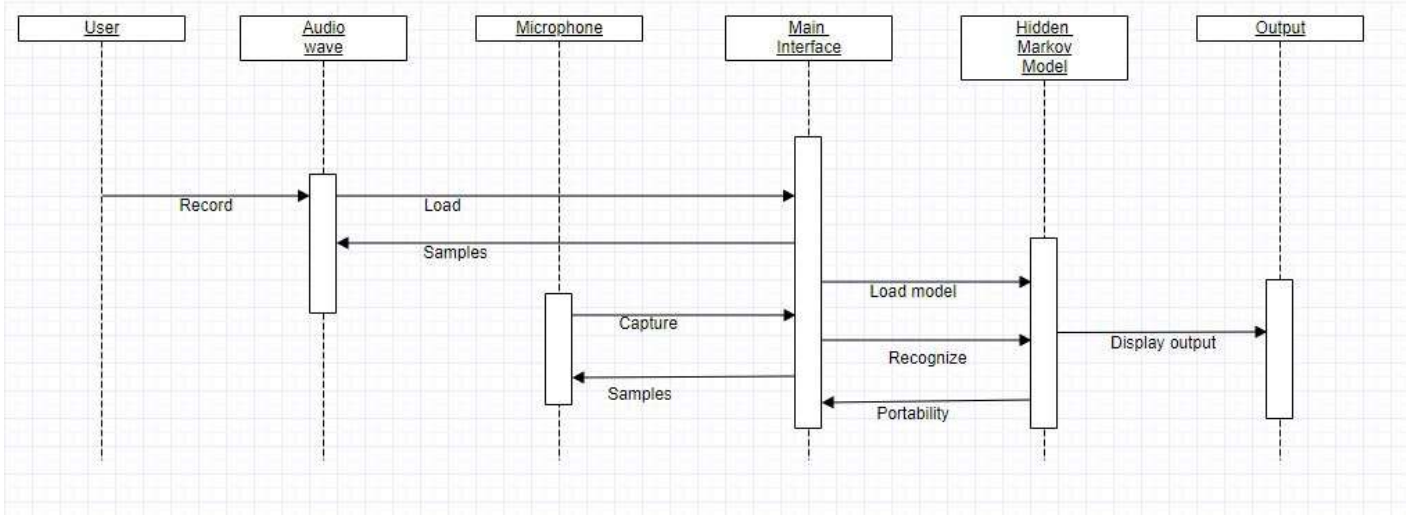

Fig 3.4 Sequence diagram

CHAPTER 4

PROJECT SPECIFIC REQUIREMENTS

Hardware: Linux or Windows operating system, RAM 4-8 GB, Microphone

Software: Python 2.7+, Anaconda or Pycharm, Ubuntu 14, CMU Sphinx, Office Tools

CHAPTER 5

IMPLEMENTATION

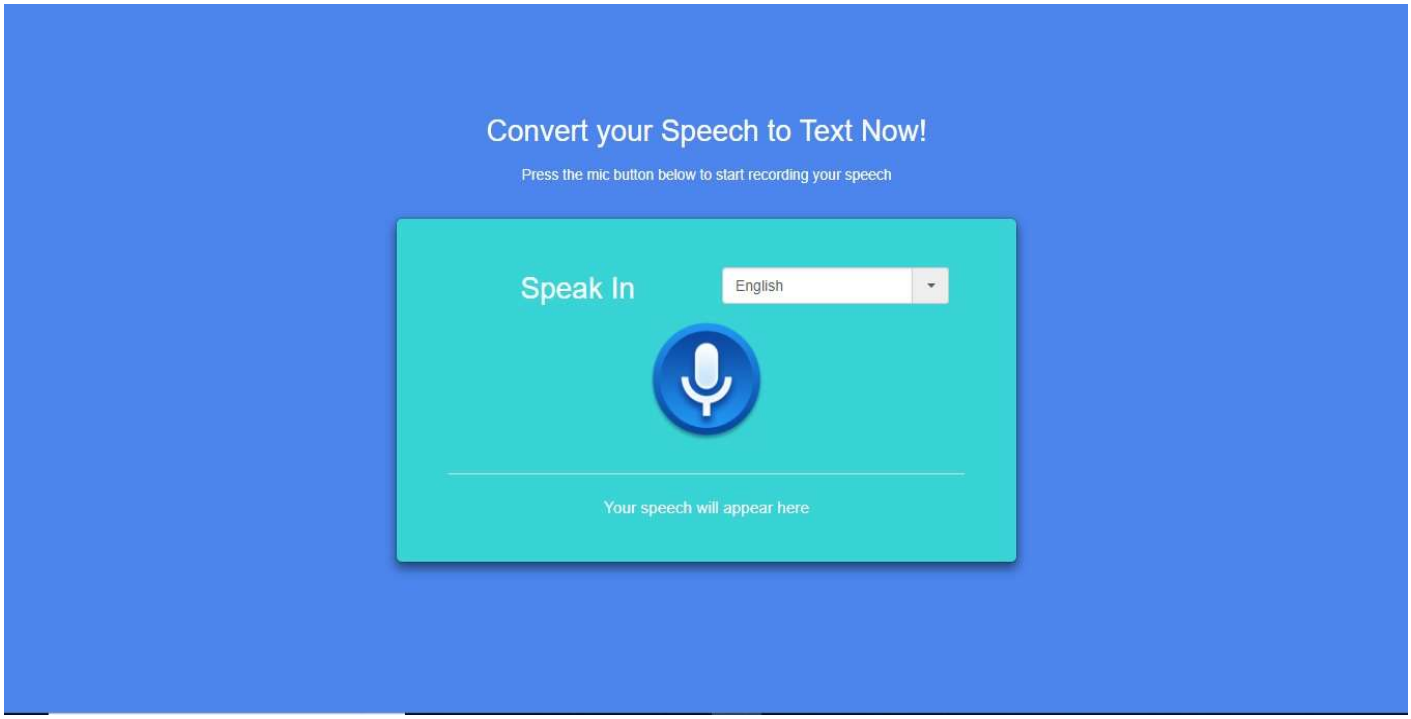

FIG 5.1 Front End

```
INFO: continuous.c(252): Ready....
INFO: continuous.c(261): Listening...
INFO: cmn_live.c(120): Update from < 41.00 -5.29 -0.12  5.09  2.48 -4.07 -1.37 -1.78 -5.08 -2.05 -6.45 -1.42  1.17 >
INFO: cmn_live.c(138): Update to   < 25.20 11.51 -8.04  4.19 -5.43 -4.97 -14.89 -7.30 -9.77 -1.03  5.96  1.42  1.08 >
INFO: ngram_search_fwdtree.c(1550):     2715 words recognized (4/fr)
INFO: ngram_search_fwdtree.c(1552):    90555 senones evaluated (149/fr)
INFO: ngram_search_fwdtree.c(1556):    44468 channels searched (73/fr), 6910 1st, 27139 last
INFO: ngram_search_fwdtree.c(1559):     3576 words for which last channels evaluated (5/fr)
INFO: ngram_search_fwdtree.c(1561):      961 candidate words for entering last phone (1/fr)
INFO: ngram_search_fwdtree.c(1564): fwdtree 2.92 CPU 0.480 xRT
INFO: ngram_search_fwdtree.c(1567): fwdtree 12.64 wall 2.075 xRT
INFO: ngram_search_fwdflat.c(302): Utterance vocabulary contains 11 words
INFO: ngram_search_fwdflat.c(948):     2573 words recognized (4/fr)
INFO: ngram_search_fwdflat.c(950):    64124 senones evaluated (105/fr)
INFO: ngram_search_fwdflat.c(952):    40811 channels searched (67/fr)
INFO: ngram_search_fwdflat.c(954):     4634 words searched (7/fr)
INFO: ngram_search_fwdflat.c(957):     1361 word transitions (2/fr)
INFO: ngram_search_fwdflat.c(960): fwdflat 1.16 CPU 0.190 xRT
INFO: ngram_search_fwdflat.c(963): fwdflat 3.33 wall 0.547 xRT
INFO: ngram_search.c(1250): lattice start node <s>.0 end node </s>.587
INFO: ngram_search.c(1276): Eliminated 2 nodes before end node
INFO: ngram_search.c(1381): Lattice has 607 nodes, 192 links
INFO: ps_lattice.c(1376): Bestpath score: -13996
INFO: ps_lattice.c(1380): Normalizer P(O) = alpha(</s>:587:607) = -781044
INFO: ps_lattice.c(1437): Joint P(O,S) = -812068 P(S|O) = -31024
INFO: ngram_search.c(872): bestpath 0.04 CPU 0.006 xRT
INFO: ngram_search.c(875): bestpath 0.13 wall 0.021 xRT
OPEN BROWSER NEW E-MAIL BROWSER WINDOW
```

FIG 5.2 Output

REFERENCES

[1] M.A.Anusuya and S.K.Katti, "Speech Recognition by Machine: A Review", (IJCSIS) International Journal Computer Science and Information Security, vol 6, no. 3, pp.181-205, 2009

[2] Rajesh Kumar Aggarwal and M. Dave, "Acoustic modeling problem for automatic speech recognition system: advances and refinements Part (Part II)", Int J Speech Technology, pp.25-32, 2012

[3] Sanjib Das, "Speech Recognition Technique: A Review", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248 9622 Vol. 2, Issue 3, May - Jun 2012