

Multilingual Speech and Text Recognition and Translation using Image

Sagar Patil¹,
Student,
Computer Department,
Rajiv Gandhi Institute of Technology,
Maharashtra, India

Mayuri Phonde²,
Student,
Computer Department,
Rajiv Gandhi Institute of Technology,
Maharashtra, India

Siddharth Prajapati³,
Student,
Computer Department,
Rajiv Gandhi Institute of Technology,
Maharashtra, India

Saranga Rane⁴,
Student,
Computer Department,
Rajiv Gandhi Institute of Technology,
Maharashtra, India

Anita Lahane⁵
Assistant Professor,
Computer Department,
Rajiv Gandhi Institute of Technology,
Maharashtra, India

Abstract - The aim of our project to automate the application to overcome from the language barrier among countries and also states within the country, the above mentioned application will perform the various features in the application. The application recognizes speech (human matter) in one language to another user defined language to communicate in expressive manner. It includes 4 modules voice recognition, translation and speech synthesis and image translation and gives audio of the translated language. Also the application accepts text written and converts it into the language needed. Application is able to recognize the text present in the image which stored in system or captured using camera and translate the text into the language needed and display the translation result back on to the screen of system.

Keywords—Speech Recognition, OCR technology, Language Translator, image extraction, text to speech

1. INTRODUCTION

Nowadays in communication the language barrier are create problem for successful communication for this we introduced this application .Speech recognition and text translation are mainly used for converting the speech to text and text to speech for understanding the language which are spoken by user during communication, because of this person can recognize the speech are spoken by other person. For image to text translation are used the OCR technology. Optical character recognition (OCR) is used to extracting the text from images which can be handwritten, signboards etc. This image extraction is used to understanding the language of sentences which into image.

1.1 problem statement

Use of mobile devices has increased a lot. Many Text to speech , speech recognition, multilingual translation, text extraction applications are developed for mobile users. But similar type of applications is not developed for desktop users. We are creating an application that consist all of the above application in one single application.

1.2 Objectives

Our main aim is to combine all different tasks such as speech recognition, text translation, text synthesis and text extraction from image all embedded in one so that we get a user friendly application

1.3. Scope

We develop this application for desktop application. Here we are integrating the speech to text, text to speech, image extraction and language translator in one system so user doesn't have to download for the different application.

2. LITERATURE REVIEW

The review [1] gives detail of speech which is recognized by system for converting into another form. It also gives the information regarding the types of speech that is continuous speech, isolated speech and Spontaneous Speech. It gives the detail of application of speech recognition of various domains. Here it also defined the approaches are used for speech recognition. The research paper [3] are representing for the text to speech converter for computer based system. In this it gives the components are used for implementing the text to speech converter The research paper [4] are examined for extraction of text from image. Here it gives the detail of phases which perform for extracting the text from image and translate them for smart phone application.

3. PROPOSED SYSTEM

The aim of the proposed system is to develop a system that has capability to perform Translation, Converting text to speech, Speech Recognition, Text Extraction. The system proposed here will be developed for a small domain of English words

4. METHODOLOGY

For our system, it has 4 modules that is text to speech, speech to text, image extraction and language translator, they are integrated are with each other.

Text to Speech

The main aim of text to speech conversion system is to convert any random or chosen text into speech. Speech can be obtained as output by concatenating recorded speech which is stored in database.

There are mainly two components, first is to processing on input text and second is to converted speech language. Normally process of conversion of text to speech is called speech synthesis and for this purpose computer system is used is called speech synthesizer. It majorly composed of two major tasks, which includes text normalization or tokenization. This process assigns phonetic transcription for each word. Divide it into phrases, clauses and sentences.

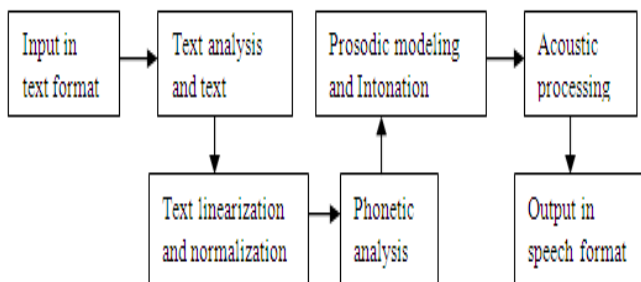


Fig. 1. Text to speech conversion process

1. Text Analysis and Detection

In text analysis it analyzes the input text and organized into list of words. And it detects the word from the database.

2. Text Linearization and Normalization

Basically text normalization is the process convert text into pronounceable form.

The text normalization process is used for conversion of uppercase and lowercase letter also it removes punctuations.

It is better used in comparison of characters which represent same meaning. Like “Can’t” and “Cannot” , “I’ve” and “I have” , “Don’t” and “Do not”.

Abbreviation conversion, Word segmentation, Number conversion, Acronym conversion are the phases of Normalization.

3. Prosodic Modeling & Intonation

Prosody is a unit of speech and it includes many features of speech other than the sound of word which is being spoken.

Prosodic analysis and its modeling include the timing of speech, pitch and rate of speech, pausing between words.

Intonation of speech includes variation in the expression when we spoke some word or sentences.

4. Phonetic Analysis

Phone is smallest unit of sound and collection of many phones grouped into phonemes. In US English it has around 45 phonemes that include vowel sound and consonant sound. It is like for example “time” contains three phonemes as “t ay m”.

5. Acoustic Processing

Finally, phonemes and prosody both are used to produce the speech waveform for each words and sentences. There are the two processing ways, first is concatenation of chunks of recorded speech. Chunk is basically grouping of words. And second process is formant synthesis using signal processing techniques.

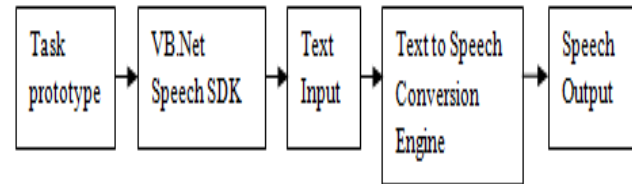


Fig.2 Block diagram of Text to Speech conversion

The VB.net Speech Software development kit is used to compile the desired program or code module. Text input block is used to feed the data text directly into editor. Synthesizer block converts the text to speech as per the input text. Speech output delivers the sound of the corresponding text into desired manner.

Text Extraction Module:

Our next module is to extract text from images. The text extracted from the images is than passed to Text to speech module if required.

In this module we have used the OpenCV.dll and Emgu.CV wrapper. The Emgu.CV is used for calling the OpenCV functions from VB.Net. EmguCV contains six dlls. Emgu.CV.dll, Emgu.CV.UI.dll, Emgu.CV.GPU.dll, Emgu.CV.ML.dll, EmguCV.OCR.dll, Emgu.CV.Util.dll. Emgu.CV.UI.dll is used for interface like image box. Emgu.CV.OCR.dll uses Tesseract-OCR library for optical character recognition. This module extracts the text from the images using the Tesseract object and displays them on the screen.

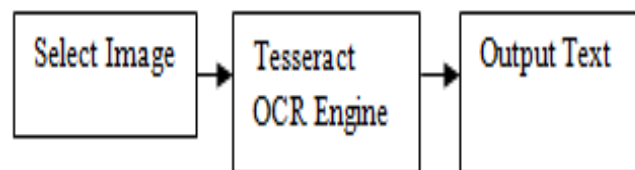


Fig. 3.Text Extraction form image

Speech Recognition:-

The speech recognition engine takes audio as input and turns into text form a gives the text as input. The speech recognition process has a front end and a back end. The front end processes the audio, isolates it into segments of sound and converts it into numeric values. This value is used to categorize the vocal sound in signal.

The back end is search engine that takes input from the front end and search them across the following databases:

The acoustic model consists of the acoustic sound which is trained to recognize the speech pattern.

The Lexicon database consists of all the words of the language it tells how to pronounce a word.

The language model is used for making proper combination of words.

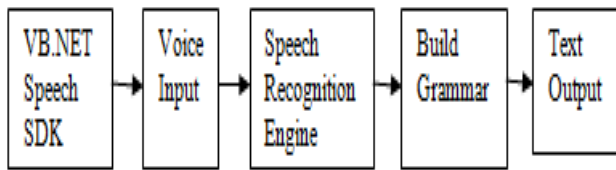


Fig. 4. The block diagram of Speech Recognition.

The VB.net Speech Software development kit is used to compile the desired program or code module. This module uses windows speech recognition. The voice is recorded using windows speech recognition. The Speech SDK has its own grammar that is used to display the recognized text as output.

Text Translation:

In Text translation we take text as input and convert it into another language. The base language is English. The text is split into words and then it is search in the dictionary and the corresponding matched text/word from the dictionary is displayed.

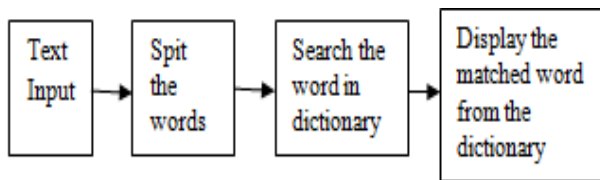


Fig. 5. Block diagram for text translation

5. FUTURE WORK

Here we implement this system for desktop application, so in future we can implement this system for mobile phone. So users can more efficiently use this system by one click of mobile instead of caring desktop for conversion of languages.

6. CONCLUSION

In this proposed system, we implemented the system for user who phasing problems of language barrier and also it user interface is also user friendly so that user can easily interact with this system .so because of this system don't have to use dictionary for understanding the meaning of word, so it automatically reduce the user task for understanding the languages for communication.

ACKNOWLEDGEMENTS

As the outset we offer our sincere thanks to our honorable guide Miss. Anita Lahane for her guidance and also encouraging us with her knowledge and experience for the development process of the project. We also value her eagerness and enthusiasm in encouraging us to develop our technical and creative ideas, which ultimately led to success of our project. Our special thanks to faculty members of Computer Engineering Department for their great support and kind co-operation to provide us with whatever we require for our project.

REFERENCES

- [1] M.A.Anusuya, S.K.Katti, "Speech Recognition by Machine: A Review", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009
- [2] Shyam Agrawal, Shweta Sinha, Pooja Singh, Jesper Olsen,"Development of text and speech database for Hindi and Indian English specific to mobile communication environment".
- [3] D.Sasirekha, E.Chandra, "Text to speech: a simple tutorial", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, And March 2012.
- [4] A. A. Tayade, Prof.R.V.Mante, Dr. P. N. Chatur,"Text Recognition and Translation Application for Smartphone