MULTILINGUAL TEXT RECOGNITION

Neeraj Yadav,
(1716110131)

Nikhil Yadav,
(1716110135)

Prakash Singh,
(1716110147)

Prabhat Vishwakarma,
(1716110145)

Mentor:-

Mr. Aditya Tandon
Asst. Professor, Department of CSE

Abstract-The aim of our project is to automate the creation of the application, in order to overcome the language barrier between the countries, but also countries as well as in the country, it is in the above-mentioned application, the performance of the various functions within the application. The app will recognize the voice work of people in one language to the other with user-defined language for communicating in an expressive way. It consists of 4 modules for speech recognition, translation, and speech synthesis, as well as to image translation, and provides the audio of the translated language. In addition, the app only accepts the written word and translates it to the language you want. The app will be able to get the text to an image that is stored in the system, it takes the camera to assist the interpretation of the text, the language and the tone of the translation is the result back to the system screen.

Keywords-Speech Recognition, optical character recognition Technology, and Language

The translator, image to delete, text-to-speech conversion

INTRODUCTION

At the moment, due to the language barrier can be a problem for successful communication, we have launched the app .Speech recognition and translation, are mainly used for converting speech to text, and text-to-speech (tts) to understand some of the language that is spoken by the user at the time of the communication, because they were able to recognize the speech, which is spoken by another person. It uses the OCR technology to translate text. Optical Character Recognition (OCR) to extract text from images, which can be handwritten, digital signage, etc., etc. This image will be used for the understanding of language and of sentences that can be converted into an image.

1.1 task

Mobile device usage has increased significantly. A lot of the text-to-speech , voice recognition, multi-language, translation, and text to the extraction of applications that are designed for mobile users. However, it is intended to be for desktop users. We are creating an app that includes all of the above-mentioned apps are in one app.

1.2 Objectives

Our main goal is to combine the variety of tasks, such as speech recognition, text translation, text synthesis, and the text extraction from an image, all of them are integrated into one single app, so that we can get in easy-to-use application.

1.3. Scope of application

We have been developing this app is to interact with the desktop app. Here, we have made the integration of speech - to-text, text - to-speech, video, and the language of extraction, translation into a single system, so that the user does not need to download any other application.

2. LITERATURE REVIEW

In the review [1], and it shows in the detail of the text of the address, which is recognized by the system, in order to convert it into another form. It also provides information on the speech, which Continue the Speech zone the Speech and Spontaneous Speech. It also explains in detail the use of speech recognition methods in the various fields. There are some of the approaches used for speech recognition, on here. The research paper [3] provides the information in a text-to-speech converter and a computer's operating system. This article discusses the components which may be used for the implementation of a text-to-speech converter in a research paper, [4] and the text on the page. Here, in detail, the steps that have to be performed to get the text from an image and translate it into a smartphone app.

3. PROPOSED SYSTEM

The objective of the proposed system is to develop a system that has the ability to perform translations, text-to-speech conversion, speech recognition, and text-to-shot. The system proposed here will be produced in a small area.

4. METHOD

It has 4 modules to our system: the text-to-speech, speech-to-text, image extraction, and file. They are integrated with each other. The text-to-speech (tts) conversion

The main purpose of a text-to-speech system is to convert any or only the selected text-to-speech. The speech can be received as an output via a recording of the voice connection, which is stored in a database.

There are basically two parts, the first of which is related to the input for the text, and the second to the output of speech to language. Usually, it is the process of converting the text into speech, it is referred to as speech synthesis, and the use of a computer system by a voice synthesizer. In short, it consists of two major tasks, such as text normalization, or tokenization. This process is the phonetic pronunciation of every word. Break it up into phrases, sentences, and any suggestions
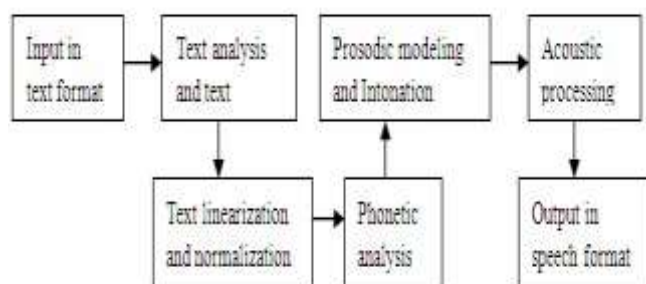
1. Text analysis and recognition

 In the analysis of the text, analyzing the input text, and to sort them into a list of words. And it does not detect a word in the database.

2. The linearization and normalization of the text

 In short, the text normalization is the process of converting the text into a specific shape and form.

The text normalization process is used to convert the uppercase and lowercase letters as well as punctuation marks.

It has been used in order to compare the characters to have the same meaning. If "Can't" and "Can't" , "I" and "me" , "No" and "No".

The combination of conversion, word segmentation, and the Number of conversions, as well as the link to convert this was the standard stage.

3. Prosodic Modeling and tuning

Prosody) a unit of speech, and many of the functions of speech, except for the sound of the voice.

Prosodic analysis and modeling of the speech time, speech pitch and rate, and pauses

in between the words.

The tone of your voice, it means a change in expression, when we say a certain word or words in the sentence.

4. Phonetic analysis

A direct dial telephone, it is the smallest unit of sound, and a collection of many of the phones are grouped into phonemes. In American English, there are about 45 phonemes, which are equipped with the sound and letters. It is as if, for example, the "period of time" consists of three phonemes, such as the

"t, a, g, m & m's.

5. Acoustic Treatment

Finally, the phonemes and prosody can be used to create a voice signal, for each word in a sentence. There are two methods of processing. The first one is a combination of the snippets of recorded speech. Chunk is basically a group of words. The second process is instagram of synthesis with the aid of signal processing techniques.
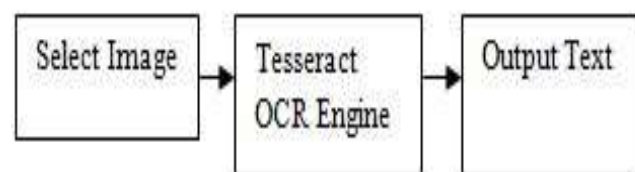
Speech recognition - -

The speech recognition engine have an audio input, and converts it into the form text and returns the text as the input. In the speech recognition process, it has a front and a back side. The interface supports audio, breaks it into segments, and then convert it to a numeric value. This value is used for the classification of vocal sounds in the output.

The back-end is a search engine, it accepts the input from the front-end, and carries out search of the following databases:

An acoustic model is composed of an acoustic sound, which is used for training a speech recognition model.

The dictionary database is made up of all of the words in the spoken language, and how the pronunciation of a given word.

The language model has been used to create just the right combination of words.

In VB.net in A speech, and a software development kit that is used in the compilation of a program, or a software module. The input block is used for the input text, and the data is sent directly to the editor. The synth is a block that converts text-to-speech, according to the text. The message to ensure that the text of the sounds in the direction you want.

The Text Extraction Module:

Our next module is used to extract text from images. If necessary, change the text, the image will be passed to the Text-to-speech module.

In this module, we have used the OpenCV.dll and, Emgu.The Wrapper off of your resume. Emgu.De the SUMMARY is to be used for the invocation of a function with OpenCV VB.Net. EmguCV has four 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 use the picture box, the interface type. Emgu.CV.OCR.dll it makes use of the Tesseract OCR library for optical character recognition. This module extracts the text from an image using Tesseract object and display them on the screen.
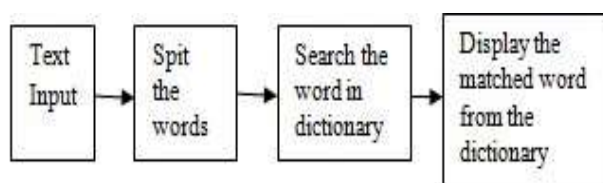

Fig. 5. Block diagram for text translation

5. WORK TO DO IN THE FUTURE

Well, here we are with the implementation of this system as a desktop application, so that in the future we will be able to implement this system in a mobile phone. In this way, the users can make use of this system to be more effective with a single click from their mobile device, rather than having to worry about the desktop, the language conversion.

6. CONCLUSIONS

This proposed system is implemented using a system, the user of the stages of the problem of the language barrier, and the interface is user-friendly so that the user may be able to work with the system .It is the result of this system is that you don't need to use a dictionary to understand the meaning of a word, that is, it automatically reduces the responsibility of the user to understand the language of communication.

ADMISSIONS

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