Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-5, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

Text To Speech Synthesizer

Surabhi Dusane, Monica Ahuja, Rucha Ghodke &
Prathamesh Kothawade
Department of Computer Engineering, K.K.Wagh Institute of Engineering Education &
Research Savitribai Phule Pune Unversity

Abstract: *Text to speech synthesizer will be a system that will combine functionality of optical recognition system i.e tesseract, machine translate system and speech synthesizer. The objective will be to develop user friendly system which will extract text from images, convert the extracted text into user friendly language and then it will convert it into audio and an animated video which describes the text more efficiently. The OCR will take image as the input and extracts text from that image, the extracted text will be given as input to machine translate system which will convert text into user friendly language. And speech synthesizer will convert it into audio. The system would be useful in various applications. It will be mainly designed for people who ca not read text or those who come across text in unfamiliar language.*

Keywords: *Optical Character Recognition, Tesseract, Machine translate, Text extraction, Text to Speech (TTS).*

1.INTRODUCTION

A) OCR: OCR stands for Optical Character Recognition. OCR is a technology that enables you to extract text from image. OCR will process images that will contain text [2]. OCR will recognize handwritten as well as printed text by an optical mechanism. The performance of OCR will depend on quality of input images. If the images will contain defect like distortion at the edges or dimmed light, it will become difficult for most OCR applications, to correctly recognize the text.

B) TERSSERACT: Tesseract has features like extensibility and flexibility. Tesseract will recognize words from images. Recognition will be carried out in two passes. In the first phase, words will be recognized, then satisfactory word will be passed as training data to Adaptive Classifier. Adaptive Classifier recognizes the text more accurately. In second phase, the words which were not recognized well in first phase will be recognized again through run over the page[3].

C] Machine Translate: Machine translation process includes translation of one natural language into another using computer software. The natural language can be any language like English, German, Marathi, Hindi etc. Even though it seems straight forward, it is complex to implement.

D] Speech Synthesis: In this phase, system can read aloud text automatically. The input text is obtained from previous phase that is language translator. It is a process to generate human speech artificially.

2. LITERATURE SURVEY

The literature survey carried by Amarjot Singh, Ketan Bacchuwar, and Akshay Bhasin was published in a Survey of OCR Applications which stated that OCR is electronic translation of handwritten, typewritten or printed text into machine translated images. OCR is widely used to recognize text from electronic documents. It uses an enhanced image segmentation algorithm based on histogram equalization using genetic algorithms for optical character recognition. The literature survey carried out states that initially Forms were scanned through an imaging scanner, faxed, or computer generated to produce the bitmap. There were two separate methods in OCR system: First text extraction and then speech translation. In a typical OCR systems an optical scanner was used to digitize input characters. Each character was then located and segmented, and then for noise reduction and normalization, resulting character image was fed into a preprocessor but during the later stages the user begins by capturing an image containing text of interest using the Mobile camera. The area containing text of the image is processed on the device to optimize it for transfer and input to the OCR.

3. PROPOSED SYSTEM

Text to speech synthesizer will extract text from image using Tesseract. It will perform language translation using Google API Also it will be using Flite algorithm for speech conversion.

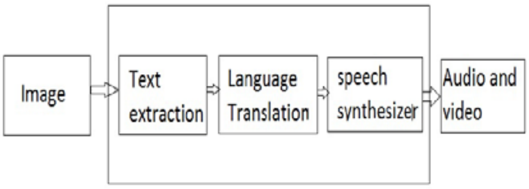

Fig 1.1: Block diagram of Text to speech synthesizer

A] Text Extraction:

Tesseract is nothing but an open source OCR engine. It was firstly released by Hp. Tesseract can work on both black text on white background and inverse of that. Tesseract works with some independent components and those components are as follows:

a] Adaptive thresholding: Adaptive thresholding is a first component of tesseract engine which takes any image as input and gives output as binary image as tesseract works on binary image. Adaptive thresholding is performed by calculating thresholding value which can be obtained from mathematical equation, thresholding value = highest pixel value + Lowest pixel value/2. After calculating thresholding value we can obtain binary image by categorizing pixels into two classes that is pixel with foreground color and background color. Pixel having lesser value than thresholding value will be categorized as pixel with background color and those having higher values will be categorized into foreground color.

b] Connected component analysis: It takes input as binary image which obtained from previous component i.e., adaptive thresholding. Connected component analysis determines the character outlines and stores it.

c]Find text lines and words: It takes input as character outlines from the previous component and the outlines are gathered together to form a blob. These blobs are then arranged to form a text line. The text line is then broken down into words with help of space between the words. The words are analysed for fixed pitch and proportional text.

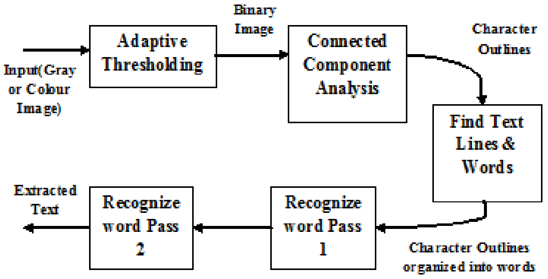

Fig 1.2. Architecture of tesseract

d]Adaptive classifier: Adaptive classifier consist of two recognition passes. The first pass tries to identify the words. The satisfactory identified words are then passed to next adaptive classifier to recognize words accurately. The words which were not determined correctly in first pass are identified accurately in this phase.

B] Language Translation:

Machine translate is used to translate the language which translates natural languages using software. The meaning of text in one natural language should not change after the translation. It is a complex process as only word to word substitution is not language translation. All the text elements are analysed and interpreted by the translator. Expertise in syntax and semantics need to be achieved so as to translate language accurately. There are three methodologies available such as Statistical , Rule based and Hybrid machine translate. We are using google API to translate language which uses Statistical machine translate.

Statistical machine translate uses bilingual text corpora which is nothing but a huge database of good translations that contains text that has been already translated into multiple languages and then uses those translated text to translate language automatically. Statistical MT method translates the language quickly but it highly depends on the text corpora that is existing. Statistical MT requires highly configured hardware to give average performance.

There are three techniques to translate language using statistical machine translate and those are word based MT, Phrase based MT and syntax based MT.

Word based MT translates language using or replacing words from text corpora whereas Phrase based MT translates sentences with the help of existing text corpora. Syntax based MT translates syntactic unit rather than translating words or string of words. There are many parallel corpora in machine-readable format and even more monolingual data. Generally, SMT systems are not tailored to any specific pair of languages.

C] Text to Speech synthesis :

In this phase, system can read aloud text automatically. The input text is obtained from previous phase that is language translator. It is a process to generate human speech artificially.

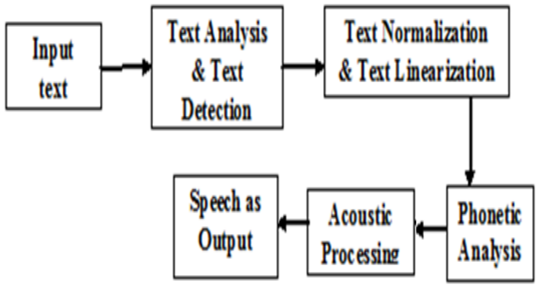
Input text
Text Analysis & Text Detection
Text Normalization & Text Linearization
Phonetic Analysis
Acoustic Processing
Speech as Output

Fig 1.3. Text to Speech synthesis

a)Text Analysis & Detection: This phase is part of preprocessing. It detects the text locations from input. It analyzes the text and prepares a list of words. Then it transforms them into full text.

b) Text Normalization & Linearization: Text normalization is process of transforming text into pronouncable form. It is performed before any text is processed for generating speech. The main objective of this process is to identify punctuation marks and pauses between words, remove punctuations, accent marks , stop words or "too common words "and other diacritics from letters.

c) Phonetic Analysis It generatess phonetic alphabets. The grapheme to phoneme conversion is done. that is nothing but a conversion of orthographical symbols into phonological symbols

d) Acoustic Processing. It performs format synthesis. It works intelligently and thus does not require any kind of database of speech samples. For speak out the text, it uses voice characteristics of a person

4. CONCLUSION

The Text to speech synthesizer system will help blind and illetrate people to understand the text and also help people to who come across the situation where they face unfamiliar language by processing the image and extract text from it using tesseract then extracted text will be passed to next module that is language translation where language of extracted text will be translated and then translated text will be passed for speech conversion. As a result people will be able to listen audio of translated language.

5. REFRENCES

[1]International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-1, March 2013 .

[2] R. Smith. "An overview of the Tesseract OCR Engine." Proc 9th Int. Conf. on Document Analysis and Recognition, IEEE, Curitiba, Brazil, Sep 2007, pp629-633.

[3]The Tesseract open source OCR engine, http://code.google.com/p/tesseract-ocr.

[4] R.W. Smith, The Extraction and Recognition of Text from Multimedia Document Images, PhD Thesis, University of Bristol, November 1987.

[5] Brown et al, "The Mathematics of Statistical Machine Translation", Computational Linguistics, 1993.

[6] Jelinek, F . "Statistical Methods for Speech Recognition" The MIT Press 1998.

[7]OmidKarami, Student of Vienna University of Technology" The brief view on Google Translate Machine. "

[8] Tejashree M. Shinde1, V. U. Deshmukh2, P. K. Kadbe3

"Text to Speech Conversion Using FLITE Algorithm."