INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 6.078
(Volume 6, Issue 3)
Available online at: www.ijariit.com

Smart Spectacles

Sai Sanketh Kulkarni
kulkarni.sanketh@gmail.com
B. V. Raju Institute of Technology, Hyderabad, Telangana

Sunnam Rajeev Sharma
rajeevsharma180104@gmail.com
B. V. Raju Institute of Technology, Hyderabad, Telangana

Y. Manish Reddy
manishreddy753@gmail.com
B. V. Raju Institute of Technology, Hyderabad, Telangana

Boddhana Chandra Shekar
bdn5171@gmail.com
B. V. Raju Institute of Technology, Hyderabad, Telangana

P. Rajesh
rajesh.p@bvrit.ac.in
B. V. Raju Institute of Technology, Hyderabad, Telangana

ABSTRACT

This paper proposes a method at developing a complete system in which Text can be converted to Speech, Text file can be converted to Speech, Text in various Languages can be converted to Speech, Image can be converted to Text and Image can be converted to Speech using PYTHON as a programming tool. The Motivation behind developing this system is to combine various modules using modular approach in order to get a simple yet effective way for differentially abled people to interact with others and thereby making the society better.

Keywords— *Text-to-Speech, Python, Raspberry Pi*

1. INTRODUCTION

Text-to-Speech module aims to provide a user- friendly application to general users. The main modules used in this application are: Text-to-Speech convertor and Image-to-Text convertor. The application provides a multi-functionality platform for users to communicate, listen or narrate conveniently. The users can choose to convert readable images to text files or read text as such. The text-to-speech mode converts a text file or inputted text to speech which then is narrated/read using the voice database used by Microsoft SAPI. The application integrates a narrator to help the users use the software. This processing is done using phonemes and concatenating syllables using optimal coupling algorithm. Image-to-text mode converts readable images to a text file which can further be used for speech conversion. Readable images are images that have less complexity in the foreground making it possible to extract the letters in a grammatical fashion. The user is provided with multiple options in this software where he/she can select the mode of operation. When Text-to-Speech mode is selected, the user has to input text using the text input box or a text file. The text is processed and the resulting speech is produced. When Image-to-Text mode is selected, the user has to input a readable image file which is processed and the text in the image is stored in a text file. The user is given an option to read this generated text file.

1.1 Motivation

To read text aloud to people who are visually impaired or legally blind.

1.2 Objective

The main objective of our present work is to provide a reliable, cost effective, low power solution for a blind people which would help them to read almost like any other normal person. The cost of this system makes it affordable for the majority of the society which in turn an effective device for them to spend on, just for once and assures wonderful travel guidance for them.

2. RELATED WORK

Currently the various methods used for text to speech conversion are Concatenation Synthesis which includes unit selection, diphone and domain specific synthesis.

Other methods include Formant, Articulatory, HMM based and sine wave synthesis. The threading of fragments of speech is done using Concatenation synthesis. This produces a natural lifelike system created speech. However, due to the differences in the nature of human voice and synthetically produced machine voice, there is a chance of recognizable glitches when producing the output. There are three key sub-types of concatenation synthesis.

Formant synthesis produces an output by using additive synthesis of an acoustic representation in the form of a model. This is different from other techniques as human speech is not used during runtime. Articulatory synthesis is a method of synthesizing machine speech based on simulations of the vocal tract of humans and their respective articulatory processes. These methods are prone to various challenges such as Text

normalization difficulties-which is the process of standardizing text and is seldom straightforward which stands as an obstacle to any speech module. Texts often contain numbers and acronyms which must be then extracted into phonetic depiction for further processing. Moreover, there are many words in the English vocabulary which contextually require a different pronunciation. There are also Text-to-phoneme challenges like determining the tone and pronunciation of a word/phrase based on the spelling in context.

Evaluation challenges occur while processing speech. Hence, there is always a compromise between the production proficiency and replay prerequisite required in speech synthesis. Prosodics and emotional content are also important for producing the vocal features that human uses showcasing their emotions and the context of the phrase/text and hence is used to produce a more natural synthesized speech. Image processing [2] is done by saturating the color of the image to grayscale using a set of grayscale thresholds. Image Segmentation is done to the noiseless grayscale image using framelets to extract the characters out of the image and the extracted characters are compared to the database and the text is produced. The size of the frames is set to identify the characters.

3. COMPONENTS USED

3.1 Hardware Components

3.1.1 Raspberry Pi 3 Model: The Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor or TV, and **uses** a standard keyboard and mouse. It is a capable little device that enables people of all ages to explore computing, and to learn how to program in languages like Scratch and Python Raspberry Pi 3 Model is an ARM based credit card sized SBC (Single Board Computer) created by Raspberry Pi Foundation. With on-board WiFi/Bluetooth support and an 64bit improved Processor

Fig. 1: Raspberry Pi 3 Model

3.1.2 Raspberry pi Camera: The Pi camera module is a portable light weight camera that supports Raspberry Pi. It communicates with Pi using the MIPI camera serial interface protocol. It is normally used in image processing, machine learning or in surveillance projects

Fig. 2: Raspberry pi Camera

3.1.3 Connecting Wires: Connecting wires provide a medium to an electrical current so that they can travel from one point on a circuit to another. In the case of computers, wires are embedded into circuit boards to carry pulses of electricity

Fig. 3: Connecting Wires

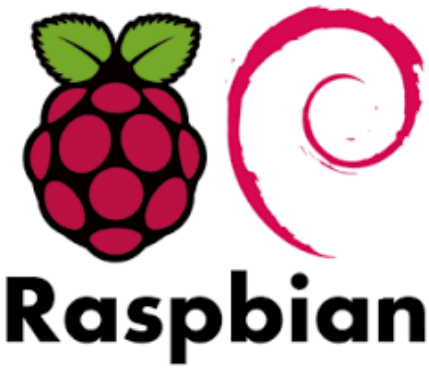
3.1.4 Spectacles

Fig. 4: Spectacles

3.2 Software Components

3.2.1 Python: Python is a general purpose and high-level programming language. You can use Python for developing desktop GUI applications, websites and web applications. Also, Python, as a high-level programming language, allows you to focus on core functionality of the application by taking care of common programming tasks.

3.2.2 Raspbian: Raspbian is the recommended operating system for normal use on a Raspberry Pi. Raspbian is a free operating system based on Debian, optimized for the Raspberry Pi hardware. Raspbian comes with over 35,000 packages: precompiled software bundled in a nice format for easy installation on your Raspberry Pi.



4. PROPOSED METHOD AND ALOGITHM

The basic block diagram project includes a modular approach in which the input text is manipulated in the text manipulation module where variations of voice, rate and volume is made.

The next block is the text to speech module where the manipulated text is converted into Speech by unit selection synthesis. The image to text module has input as image which is converted to text. Using the above two modules combined we can have image to speech module to give output as speech.

The flow of the text to speech module which is explained in detail as follows: Text to speech conversion can be accomplished by starting with the method of pre-processing of the input text. Here the text abbreviations, acronyms and numbers are expanded [10]. The pre-processed text will then be of this converted to Unicode. Unicode has the explicit aim of transcending the limitations of trade encodings. Here the pre-processed text is used to identify the fonts of input text and is converted to Unicode. Now, the encoded text is segmented into syllables and the duplicates are removed. The syllabled text is then mapped with the syllable sound files in the database. These syllables will be then concatenated and smoothened for resultant outputs. This is done by optimal coupling algorithm. This gives a smooth human speech output. Variations can be applied to the resulting output. Voice, rate, volume, of the output speech can be suitably changed by the user.

The program starts with the introduction page accompanied with a narrated welcome. On clicking the next button, the user is directed to the help page where the user gets the option to invoke the narrator which reads out the contents of the help page by click the read button. By clicking onto the enter button, the user is introduced to a dialogue box containing options for manipulating voice, rate, pace of the text to be read. The user will be directed to page of his choice of function where along with application of the function, the user can also control the volume of the speech output. The user enters the text into the text field and the resultant speech output is produced upon clicking on the speak button. The Voice function allows the user to choose bet the language voice packs installed on the system.

Hence, the function allows the provision of generating a multi-lingual output. The Rate function provides the provision to change the sampling rate of the Speech output. The Pace function allows the user to control the speed of the resultant speech output. After completion of the various processes, the user gets the option between exiting the app or reusing it to operate other functions.

4.1 Applications

Developed to aid the visually impaired by offering a computer-generated spoken voice that would “read” **text** to the user.

5. CONCLUSION

The paper objective is underlined by the necessity of voice assistant system for the increasing number of blind people all over the globe. The "Smart Glass for Blind People" is practically, a feasible device and can be conveniently carried by any blind person. It does have a few limitations which we’ll target to solve in its future developments

6. FUTURE SCOPE

- Less cost,
- More accurate.

7. REFERENCES

[1] https://www.pyimagesearch.com/2018/09/17/opencv-ocr-and-text-recognition-with-tesseract/
[2] Google Vision API in Python by jie jenn

BIOGRAPHY

P. Rajesh
Assistant Professor
Electronics and Communication Engineering
B. V. Raju Institute of Technology, Hyderabad, Telangana, India

Sai Sanketh Kulkarni
Student
Electronics and Communication Engineering
B. V. Raju Institute of Technology, Hyderabad, Telangana, India

Sunnam Rajeev Sharma
Student
Electronics and Communication Engineering
B. V. Raju Institute of Technology, Hyderabad, Telangana, India

Boddhana Chadra Shekar
Student
Electronics and Communication Engineering
B. V. Raju Institute of Technology, Hyderabad, Telangana, India

Y. Manish Reddy
Student
Electronics and Communication Engineering
B. V. Raju Institute of Technology, Hyderabad, Telangana, India

