iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** VIII **Month of publication:** August 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.46434

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue VIII Aug 2022- Available at www.ijraset.com*

# Analog Driven Robot using IoT and Machine Learning

Yash Pradhan[1], Sonali Choudhary[2]
[1]*Student, Xavier Institute of Engineering,* [2] *Student, Thadomal Shahani College of Engineering*

***Abstract*: *Visual impairment is one of the biggest limitations for humanity, especially in this day and age when information is communicated a lot by text messages (electronic and paper based) rather than voice. Facial recognition is category of biometric software that maps an individual's facial features mathematically and stores the data as a sprint The software uses deep learning algorithms to compare a live capture or digital image to be stored face print in order to verify an individual's identity vbnnn This project aims to develop a device to help people with visual impairment. In this project, we manufactured a device that converts an image's text to speech. The basic outline is an embedded system that captures an image, extracts only the area of interest (i.e. region of the image that comprises text) and changes that text to speech. It is incorporated using a Raspberry Pi 3 and a Raspberry Pi camera module. We have two phases in our project that are Text to Speech and Facial Recognition. Every Module for picture handling and voice preparing are available in the device. It likewise can play and stop the output while reading. The expectation is that it has less error rate and less processing time and less cost productivity.***
***Keywords: Machine Learning, Computer Vision, Raspberry Pi-3, Pytesseract, GTTS, OpenCV***

## I. INTRODUCTION

In our planet of 7.4 billion people, 285 million are outwardly weakened out of whom 39 million individuals are totally visually impaired, for example have no vision by any stretch of the imagination, and 246 million have gentle or serious visual weakness (WHO).[1]

It has been anticipated that by 2020, these numbers will reach 75 million visually impaired and 200 million individuals with visual weakness. As reading is the prime significance in the day by day schedule (content being available wherever from books, business items, sign-loads up, computerized screens, travel tickets and so on.) Of humankind, outwardly disabled individuals face a lot of troubles .[1]

Individuals who experience the ill effects of low vision, sight and visual impedance are not ready to see words and letters in standard newsprint, books and magazines plainly. This can be troublesome for the person and may also moderate down the person's confidence level. This may also lower their intelligence level. Thus, a device is expected to enable them to read. So, this device can help them scan and read any sort of content by transforming it into voice messages. The motivation behind this device is to process the important documents, Reading material, and News Papers as contribution to a voice as output. Every Module for picture handling and voice preparing are available in the device. It likewise can play and stop the output while reading. The expectation is that it has less error rate and less processing time and less cost productivity. To build up the device Raspberry-Pi 3 along with its camera module and also Pytesseract (Python library) is used. It may or may not require any human supervision. The target of Text to Speech is to change over a given input into a verbally expressed waveform. Text processing and speech generation are two primary parts of a TTS system. The goal is to process the given information content and produce proper grouping of phonemic units. These phonemic units are acknowledged by the speech generation segment either by speech synthesize from parameters or by choice of a unit from a massive speech corpus. For normal sounding speech conclusion, it is basic that the text preparing segment produces a proper grouping of phonemic units relating to an arbitrary information content. Any Text-to-speech framework comprises two significant components. Beginning with the output, we need some kind of sound-creating instrument. A text to speech framework the info content is first investigated, standardized and translated into a phonetic or some other semantic portrayal. And then we have Text preparing parts which manage low level handling issues, for example, sentence division and word division [2].

It joins the idea of optical character acknowledgment (OCR) and Text to speech synthesizer (TTS) in Raspberry-Pi 3. This framework is used to assist the visually impaired individuals with interacting with computers adequately through vocal interface. Text extraction from shading pictures is a troublesome task for computers.

©IJRASET: All Rights are Reserved 

The text to speech change framework is perused the English letters in order and numbers that are in the picture utilizing the OCR strategy and converting it into the voice design. This paper displays the plan usage and trial aftereffect of the gadget.

This device comprises two sections, picture handling module and voice preparing module. The optical character acknowledgment (OCR) is the procedure that changes over the output or printed text pictures into the text arrangement for further preparation. This paper has exhibited the straight-

forward methodology for text extraction and its change into speech.

The testing of the device was done on Raspberry-Pi 3 module. Text to Speech (TTS) framework creates a more characteristic voice that can be firmly coordinated with human voice. The instances of the speech combination are the voice empowered email and informing. The initial step of speech conclusion is for the utilizations to talk a world in receivers and later that speech is changed over into the advanced organization by utilizing simple to computerized converters and concluding it to memory. Programming forms the information picture and changed over into text arrangement. Take Picture: This Square will examine the picture/page that should be heard in the sound arrangement. Change picture into dim scale/decide ROI: This will make the picture clear for the program and distinguish the districts of intrigue. OCR handling: Here the preparation will be done as such as to peruse out the content organization from or the characters that are to be heard. It will get to the PIL (Python Image Library) and gTTS (Google text to speech) and along these lines giving the sound as output. How Text-to Speech Can Helps Students: Improves word acknowledgment. Builds the capacity to focus and recall data while perusing. Permits children to concentrate on cognizance as opposed to sounding out words. Builds children's fortitude for understanding assignments. Assists kids with perceiving and fixing mistakes in their own composition.

Thus, this paper will try to overcome some of the above mentioned problems. Visual impedance is perhaps the greatest impediment for humankind, particularly at this generation, when data is imparted a ton by instant messages (electronic and paper based) as opposed to voice. Numerous difficulties are faced by a visually impaired individual in his/her everyday life while communicating. The device has been proposed to help people with visual disability. Right now, I built up a device that changes a picture's content to speech. The output will be in sound configuration. This will help people with visual weakness.

## II. WORKING

User Creation for Facial Recognition : The first step where the user will have interaction with the system through camera. Here we first try to feed in the user id or try to create a user for the system, which means the user's face will be scanned here. 30 images of the user will be clicked in this step.

Training algorithm : In this particular step the captured 30 images of the user will be used for training purposes, to know the size of the face, the dimensions, the features with the help of libraries.

Face Detection : Face detection is done basically in several shots, that is the camera module has first captured images(30) of the face in front of it. This is converted into grayscale to determine the proper tone and features of the particular face. The system will align the faces on each image so that now the analyzed part of the face is done with accuracy. Image from dataset and array : The names from the array and images from dataset are compared so as to identify the name of the face in front of the camera.

OCR : Optical Character Recognition is a particular name from the array to be read by the algorithm so as for it to be read out further.

Name Detected : The particular name and image will be read and it is ready to be spoken out in the next step. Output spoken in audio form : At this time the audio will finally be heard in the earpiece or the speaker, the calling out options can be "name" or "unknown" or "no one in frame" considering the respective situations.

### *A. Pseudo Code*

1) *Start:* This is the initial stage or rather stop of the program, it's the very first step in the facial recognition process, we initiate the program at this point.
2) *Database Entry via User ID:* This is the stage one of the process wherein, we try to create a dataset. The dataset is created in order to save the faces in the memory.
3) *Face Recognition for Images:* The faces will be recognized with the help of dimensions given in the code. There will be a minimum size of the image to be accepted and the maximum size. The recognition will be done with the help of Rectangle formation around the face.

4) *Storing The Captured Data In The Database:* The camera will be capturing images. There will be 30 images clicked at the stage one. Wherein these images are stored in a dataset as they will be forming a database for the system. The program will escape the loop only after it has captured 30 images.
5) *Training the Model:* Training of the model is required in order for it to do the machine learning part of the method. This is the stage wherein the 30 images captured in the stage one will be examined and studied by the program with the help of libraries. This is stage two of the process.
6) *Face Detection:* Now the Face recognition part done for the images previously and the training part wherein the 30 images were captured will be studied in this stage. This is where it will go into a loop and recognize the face from the array.
7) *Text to Speech:* The recognized face will be said out in the audio from on the earpiece, which is done through Optical Character Recognition (OCR).
8) *Distance Measured in Ultrasonic Sensor:* The distance at which the person is will be detected with the help of the Ultrasonic Sensor. This will be done in order to ensure the security measures for the visually impaired.
9) *End:* This where it has recognized the person in the frame and it will exit the program.

*B. Results and Discussion*

1) The output observed for image text to speech conversion is – Identify text on the image and convert it into sound document. It can change over both capitals just as little letters. The captured image is converted to the text and saved at the same location of the image. It takes approximately 7-8 sec to convert the text. The converted text is processed with the Festival or Py-tesseract. The speech is obtained as an output in Headphones. It recognizes numbers as well as letters in English Range of reading Distance is 30cm.
2) The output observed for Face Recognition is – Better Security –Additionally, appreciate better security with a facial biometrics' framework. Not exclusively would be able to follow representatives, yet any guests can be added to the framework and followed all through the territory as well. Anybody that isn't in the framework won't be given access. Simple Integration – Biometric facial frameworks are likewise simple to incorporate into your organization High Success Rate –Facial biometrics innovation today has a high achievement rate, particularly with the rise of 3d face acknowledgment. It is amazingly hard to trick the framework, so the person can have a sense of safety realizing that your framework will be fruitful at the following time also, participation while giving better security.

## III. CONCLUSION

Text to Speech can change the content of a picture into sound. For this transformation does require an internet connection. It is exceptionally simple to utilize, so the visually impaired individual can autonomously utilize this gadget. Through the usage of this project, visually disabled people would easily be able to tune in to the content of the record. Through interpretation apparatuses, one believes the content to the ideal language and afterward again by utilizing the Google text to speech he/she can change over that changed content into sound. We can likewise expand the extension for significant distance catching. The catch work could likewise be concentrated on building up an effective compact. We actualized a biometric framework utilizing a Raspberry Pi 4 and internet connection. Pictures are caught dependent on motion discovery and sent to the server where they are subjected to a face acknowledgment technique Our experiments suggest that the created acknowledgment methodology is able to accomplish a 95.

## IV. REVIEW OF LITERATURE

This chapter gives us an outline about all the related fields and knowledge for completion of this project. To get a picture of what the project is about, literature review/survey for an associate journal was done. Vinaya Phutak, Richa Kamble, Sharmila Gore, Minal Alave, R.R.Kulkarni " presented "Text to Speech conversion using Raspberry Pi". The following paper emphasizes that Text to Speech (TTS) is a kind of speech blend application that is utilized to make an expressed sound variant of the content in a PC archive, for example, an assistance record or a Web page. TTS can empower the reading of computer show data for the outwardly tested individual, or may essentially be utilized to increase the reading of an instant message. Current TTS applications incorporate voice empowered email and spoken prompts in voice reaction frameworks. TTS is frequently utilized with voice acknowledgment programs. Like different modules the procedure has got its own significance on being interfaced with, where Raspberry Pi discovers its own tasks dependent on picture preparing plans. So once a picture gets changed over to content and thereby it could be changed over from content to speech. Character acknowledgment process closes with the transformation of content to discourse and it could be applied at anyplace. [2]

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Miss.Prachi Khilari, Prof.Bhope V.P presented “A Review on speech to text conversion methods' '. The following paper emphasizes that Text to Speech converters convert’s ordinary language content into speech. Incorporated speech can be made by linking bits of recorded speech that are put away in a database. System contrast in the size of the put away speech units; a framework that stores phones gives the biggest output run, yet may need clearness. For specific utilization spaces, the capacity of whole words or sentences takes into consideration high- caliber output. On the other hand, a synthesizer can consolidate a model of the vocal tract and other human voice attributes to make a totally ”manufactured” voice[3]

Akshay.A, Amrith.NP, Dwishanth.P and Rekha.V presented “A Survey on text to speech conversion”. The following paper emphasizes that Digital Image is the utilization of computer calculations to perform picture handling on computerized pictures. As a subcategory or on the other hand a field of computerized signal preparation, advanced picture handling has numerous points of interest over simple picture handling. It permits a lot more algorithms to be applied to the information. What's more, can maintain a strategic distance from issues, for example, the development of buildup noise and signal distortion during preparation. Since pictures are characterized by more than two measurements (maybe progressively) , computerized picture preparation might be displayed as multidimensional frameworks. It is among quickly developing advancements today, with its applications in different parts of a business. Image Processing structures center research engineering and computer science discipline as well. Image handling is done on computerized pictures by utilizing computer calculations. [4] Sujata G.Bhele, V.H.Mankar presented “A Review on Face Recognition and facial expression recognition for blind people”. The following paper emphasizes that the Face recognition system has two principle assignments: confirmation and identification proof. Face check implies a 1:1 match that compares face pictures against layout face pictures whose identity is guaranteed. Face recognizable proof methods a 1:N issue that analyzes an inquiry face picture against all image formats in a face database. Machine acknowledgment of faces is getting significant because of its wide range of business and law implementation applications, which incorporate legal recognizable proof, get to control, outskirt reconnaissance and human communications and accessibility of minimal effort recording gadgets. Different biometric highlights can be utilized with the end goal of human acknowledgment like unique mark, palm print, hand geometry, iris, face, discourse, steps, signature and so on. The issue with unique mark, iris palm prints, Speech, gaits are they need dynamic co-activity of an individual while face acknowledgment is a procedure that doesn't require dynamic co-activity of an individual so without teaching the individual can perceive the individual. So face acknowledgment is significantly more beneficial different biometrics.[5]

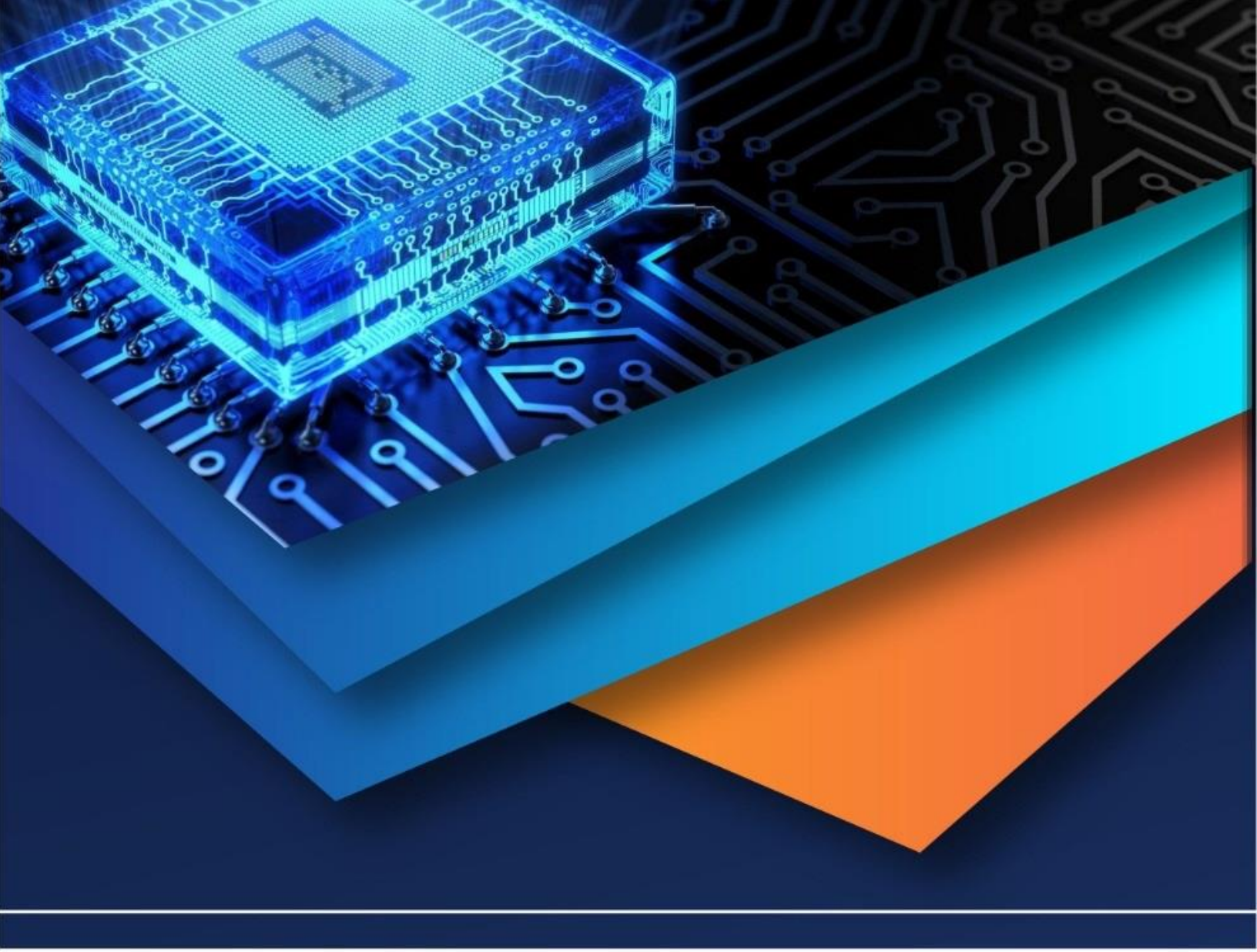
Bhupendra Vishwakar, Pooja Dange, Abhijeet Chavan, Ak shay Chava presented “Face and facial recognition”. The following paper emphasizes that An individual is recognized by his/her face. Face being the most significant part, utilized for recognizing an individual from another. Each face has various highlights and has various qualities of its own. In this way, face acknowledgment assumes a crucial job in human conduct. Specifically, outward appearances assume a significant job in the human to human correspondences and gives exceptionally solid prompt in estimating levels of enthusiasm of an individual while collaborating with a machine. In this paper ,the visually impaired will act naturally ready to recognize individuals because of face acknowledgment and will get a sound message about the individual, ”This is someone or other individual” and the visually impaired can act naturally ready to address them without hanging tight for an individual from inverse to come to him and address him, just he needs to distinguish the individual (given the individual subtleties spared in the framework database). The new faces can likewise be added to the database.[6] Aishwarya Admane, Afrin Sheikh, Sneha Paunikar, Shruti Jawade, Shubhangi Wadbude, Prof. M. J. Sawarkar presented “A Review on Different Face Recognition Techniques.”. The following paper emphasizes that Face acknowledgment is a fundamental bit of the capacity of human insight system and is a norm task for individuals, while building a practically identical computational model of face acknowledgment. The computational model adds to theoretical bits of information just as to various rational applications like motorized gathering perception, get the chance to control, framework of human PC interface (HCI), content-based picture database organization, criminal distinctive confirmation and so forth. The soonest work on face acknowledgment can be followed back in any occasion to the 1950s in mind science and to the 1960s in the planning composing. A part of the soonest inspects fuse chips away at facial hair emotions by Darwin. Regardless, ask about customized machine acknowledgment of countenances started in the 1970s.[7] Neeraj Pratap, Shwetank Arya, Nishant Rathi presented “Significance of spectral curve in Face Recognition”. The following paper emphasizes that Face Recognition is a noteworthy research issue covering numerous fields and trains. This since Face Recognition, in addition to having successive down to earth applications, for example, bankcard recognizable proof, get to control, Mugshots looking, security observing, and reconnaissance framework, is a major human conduct that is fundamental for viable interchanges and connections among individuals. Progress has progressed to the point that FR frameworks are being set up in genuine settings. The quick advancement of FR is because of a blend of elements: dynamic improvement of calculations, the accessibility of enormous databases of facial pictures, and a strategy for assessing the exhibition of

face acknowledgment calculations. FR is a biometric approach that utilizes robotized strategies to check or perceive the personality of a living individual dependent on his/her physiological qualities. All in all, a biometric recognizable proof framework utilizes either physiological attributes or standards of conduct to recognize an individual. On account of human natural defense of his/her eyes, a few people are hesitant to utilize eye recognizable proof frameworks. FR has the advantage of being an aloof, non-meddlesome framework to check individual personality in a ”common” and neighborly way.[8] Xvan Kan , Andrew Markham, Niki Trogoni presented “Autonomous learning for Face Recognition in The Wild via Ambient wireless cues". The following paper emphasizes that Facial acknowledgment is a key empowering segment for developing Internet of Things (IoT) administrations, for example, keen homes or responsive workplaces. Using profound neural systems, facial acknowledgment has accomplished astounding execution. Be that as it may, this is just conceivably at the point when prepared with several pictures of every client in various reviews and lighting conditions. Obviously, this degree of exertion

Enlistment and marking is inconceivable for wide- spread organization and reception. Motivated by the way that a great many people convey brilliant wireless gadgets with them, for example cell phones, we propose to utilize this remote identifier as a supervisory mark. This permits us to clergyman a dataset of facial pictures that are extraordinary to a specific space for example a set of individuals in a specific office. This custom corpus would then be able to be utilized to fine tune existing pre-prepared models for example FaceNet. Be that as it may, due to the ideas of remote engendering in structures, the supervisory names are boisterous and frail. We propose a novel procedure, AutoTune, which learns and refines the relationship between a face and wireless identifier after some time, by expanding the between group partition what’s more, limiting the intra-group separation. Through broad experiments with numerous clients on two destinations, we exhibit the capacity of AutoTune to plan a situation explicit, persistently developing facial acknowledgment framework with completely no client exertion. [9]

## REFERENCES

[1] Vinaya Phutak, Richa Kamble, Sharmila Gore, Minal Alave, R.R.Kulkarni”, “Text to Speech conversion using raspberry pi”, International Journal of Innovative Science Research, Volume 4, Issue 2 , Feb-2019.
[2] “Miss.Prachi Khilari, Prof.Bhope V.P”, ”A Review on speech to text conversion methods”, International Journal of Advanced Research in Computer Engineering Technology, Volume 4,Issue 7, July 2015.
[3] “Akshay.A, Amrith.NP, Dwishanth.P and Rekha.V”, “A Survey on text to speech conversion”, International Journal of Trend in Research Development, Volume 5, Issue 2, March April 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)