

ANNA UNIVERSITY: CHENNAI-600 025

BONAFIDE CERTIFICATE

Certified that this project report “**DEVELOPMENT OF AN AUTOMATED TEXT TO SPEECH DEVICE FOR VISUALLY IMPAIRED**” is the bonafide work of,

L. KINGSLY KIRUBAKARAN	-	953219106014
M. KISHORE ANTON	-	953219106015
T. SIVA KUMAR	-	953219106035
M. SIVANRAJ	-	953219106036

who carried out this project under my supervision.

SIGNATURE

Dr. K. ESAKKI MUTHU M.E., Ph.D.,

HEAD OF THE DEPARTMENT

Department of Electronics and
Communication Engineering,
University VOC College of Engineering.
Thoothukudi-628 008

SIGNATURE

Dr. A. MOOKAMBIGA M.TECH., Ph.D.,

SUPERVISOR

Department of Electronics and
Communication Engineering,
University VOC College of Engineering,
Thoothukudi-628 008

Submitted for the project viva-voce examination held on _____

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

It is the difficult task in life to choose words to express words to express one's gratitude towards in the beneficiaries. We are very much grateful to the **ALMIGHTY GOD** who helped us the way throughout the project and who has molded us into what we are today.

We wish to express our heartfelt regards and sincere thanks to our beloved Dean **Dr. C. PETER DEVADOSS M.E., Ph.D.**, University VOC College of Engineering for his generous guidance, help and constant encouragement during the course of this project work.

We wish to express our sincere thanks to our Head of the Department **Dr. K. ESAKKI MUTHU M.E., Ph.D.**, Department of Electronics and Communication Engineering, for his insightful comments and constructive suggestions to improve the quality of this project.

We have immense pleasure on expressing our sincere thanks to our Project Coordinator and our respected guide **Dr. A. MOOKAMBIGA M.TECH., Ph.D.**, Assistant Professor, Department of Electronics and Communication Engineering, for her stimulating guidance, continuous encouragement and supervision throughout the course of this work.

We wish to express our deep sense of gratitude to our parents, friends and all teaching and non-teaching staff members and all who participated enthusiastically with their constructive criticism either directly or indirectly in making this project a grand success.

ABSTRACT

This project involves the development of a text-to-speech system using a Raspberry Pi 4, a Pi Camera module 3, and Google Cloud services. The system is designed to capture an image containing text, extract the text using OCR technology, translate the text to a desired language using the Google Translate, and generate speech from the translated text using the Google Cloud Text-to-Speech API. The project aims to address the issue of language barriers, allowing users to capture images containing text in any language and receive spoken translations in their desired language. The project's use of OCR technology and cloud-based translation services ensures high accuracy and quality in the translation process. The Google Cloud Text-to-Speech API's lifelike intonation, rhythm, and pacing create natural-sounding speech, making the system highly effective in facilitating communication. The project's efficient and user-friendly design makes it an ideal tool for various applications, including education, language learning, and accessibility for individuals with visual impairments. The project's use of a Raspberry Pi 4 and Pi Camera module 3 makes it portable and accessible for users in various settings. Overall, this project demonstrates the significant impact that technology can have on breaking down language barriers and facilitating communication and access to information for individuals worldwide.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO
	ABSTRACT	iii
	LIST OF TABLES	vi
	LIST OF FIGURES	vii
	LIST OF ABBREVIATIONS	x
1.	INTRODUCTION	1
	1.1 GENERAL	1
	1.2 OCR	1
	1.3 GOOGLE CLOUD VISION API	2
	1.4 GOOGLE TRANSLATE	3
	1.5 GOOGLE CLOUD TEXT TO SPEECH API	4
	1.6 NEED FOR THIS PROJECT	5
	1.7 OBJECTIVES	5
	1.8 ORGANISATION OF CHAPTERS	6
2.	LITERATURE SURVEY	7
	2.1 GENERAL	7
	2.2 REVIEW OF LITERATURE SURVEY	7
	2.3 SUMMARY OF REVIEW	13

3.	METHODOLOGY	14
	3.1 GENERAL	14
	3.2 PRICIPLE OF THE SYSTEM BUILT	14
	3.3 GOOGLE CLOUD VISION API	16
	3.3.1 Pre-Processing	17
	3.3.2 Text Detection	25
	3.4 GOOGLE TRANSLATE	31
	3.5 GOOGLE CLOUD TEXT-TO-SPEECH	33
	3.7 DEVICE IMPLEMENTATION	36
	3.7.1 Software	36
	3.7.1 Hardware	45
4.	RESULTS AND DISCUSSIONS	51
	4.1 GENERAL	51
	4.2 SIMULATION OUTPUT	52
5.	CONCLUSION	59
	5.1 CONCLUSION	59
	5.2 FUTURE SCOPE	60
	ANNEXURE	
6.	REFERENCES	

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
3.1	Features of Google Cloud TTS	34
4.1	Accuracy for Google Cloud Vision API	57
4.2	Accuracy for Google Translate	57

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
3.1	Flow chart	15
3.2	Flow diagram of Vision API	16
3.3	Image resizing	18
3.4	Image rotation	19
3.5	Image cropping	21
3.6	Image normalization	22
3.7	Color filtering	23
3.8	Image compression block diagram	24
3.9	General OCR model	26
3.10	Image having text information	26
3.11	Extracting character boundaries	27
3.12	Naming Convention followed (Labelling)	27
3.13	ConvNet Model	28
3.14	OCR flow diagram	29
3.15	Obtaining contour for given dataset	30
3.16	Text detection using Vision API	30
3.17	Google translate block diagram	32

3.18	Google Cloud TTS block diagram	34
3.19	Activating service account key	39
3.20	Camera capturing the image	40
3.21	Extracted text output	41
3.22	Translated text output	42
3.23	Generated speech output	43
3.24	Playing speech output	44
3.25	Block diagram of the device	45
3.26	Raspberry Pi 4 Model B	46
3.27	Pi camera module V3	48
3.28	Device implementation	50
4.1	Input image1	52
4.2	Pre-processed image1	52
4.3	Boundary detected for image1	52
4.4	Translated output for image1	52
4.5	Output text for image1	53
4.6	Input image2	53
4.7	Pre-processed image2	53
4.8	Boundary detected for image2	54

4.9	Translated output for image2	54
4.10	Output text for image2	54
4.11	Input image3	55
4.12	Pre-processed image3	55
4.13	Boundary detected for image3	55
4.14	Translated output for image3	55
4.15	Output text for image3	56
4.16	Input image4	56
4.17	Pre-processed image4	56
4.18	Boundary detected for image4	57
4.19	Translated output for image4	57
4.20	Output text for image4	57
4.21	Accuracy for Text to Speech Output	58

LIST OF ABBREVIATIONS

API – Application Programming Interface

ASCII – American Standard Code for Information Interchange

BMP – Bitmap Image file

CNN – Convolutional Neural Network

ConvNet – Convolutional Network

JPEG – Joint Photographic Experts Group

ML – Machine Learning

NMT – Neural Machine Translation

OCR – Optical Character Recognition

OpenCV – Open Source Computer Vision Library

PNG – Portable Network Graphics

RNN – Recurrent Neural Network

RGB – Red Green Blue

SSML – Speech Synthesis Markup Language

TTS – Text to Speech

VS Code – Visual Studio Code

YOLO – You Only Look Once