**SILENT SPEECH RECOGNIZER AS LIP READING**

**23VAC04 EXPLORING GENERATIVE AI TOOLS**

***Submitted by***

**SAMSON DANIEL R (2012027)**

**SRI MURUGAN M (2012051)  
SHENBAGA RAJ S (2012088)**

***In partial fulfilment for the award of the degree***

***of***

**BACHELOR OF ENGINEERING**

***in***

**COMPUTER SCIENCE AND ENGINEERING**



**NATIONAL ENGINEERING COLLEGE**

**(An Autonomous Institution, Affiliated to Anna University – Chennai)**

**K.R. NAGAR, KOVILPATTI – 628 503**

**APRIL 2024**

**BONAFIDE CERTIFICATE**

This is to Certify that this project report entitled **“SILENT SPEECH RECOGNIZER AS LIP READING”** is the bonafide work of **SAMSON DANIEL R (2012027), SRI MURUGAN M (2012051), SHENBAGA RAJ S (2012088)** who carried out the project work under my supervision.

**SIGNATURE**

**Dr. V. GOMATHI. M.Tech., Ph.D.,**

**Professor & Head,**

Department of CSE,

National Engineering College

(An Autonomous Institution),

K.R.Nagar, Kovilpatti - 628503.

**SIGNATURE**

**Ms.A.Lincy,M.E.,**

**Assistant Professor,**

Department of CSE,

National Engineering College

(An Autonomous Institution),

K.R.Nagar, Kovilpatti - 628503.

Submitted to the **23VAC04 EXPLORING GENERATIVE AI TOOLS** Assessment Examination held at **National Engineering College, K.R.NAGAR, KOVILPATTI**.

**ACKNOWLEDGEMENT**

First and foremost, we would like to thank God Almighty for showering his blessings throughout our life. He has been the tower of our strength in each step of our work. We take the privilege to express hearty thanks to our parents for their valuable support and effort to complete the project work.

We would like to express our deep sense of gratitude and respectful regards to our director **Dr. S. Shanmugavel, B.Sc., D.M.I.T., Ph.D.,** for giving an opportunity to do this work.

We have great pleasure in acknowledging our Principal **Dr.K.Kalidasa Murugavel, M.E., Ph.D.,** for extending his full support to undergo this work.

We express our profound thanks to our beloved Head of the Department **Dr. V. Gomathi, M.Tech., Ph.D.,** for extending her full support and providing various facilities during the project work.

We express our gratitude to our Course Instructor, **Mr.V.Manikandaraman**, Adjunct Faculty/CSE, Director - Enterprise Automation, Global Delivery Head for Life Sciences, Cognizant Technology Solutions Engineering whose valuable guidance, technical support and suggestions helped us doing the project work his valuable guidance at each and every stage of the work.

We would like to thank our coordinators **Ms.A.Lincy, M.E.,** Assistant Professor, and **Mr.J.Karthikeyan, M.E.,** Assistant Professor, Department of Computer Science and Engineering for their valuable guidance at each and every stage of the project.

We extend our hearty thanks to our tutors and class in-charges for their valuable guidance. We are grateful to all the staff members and our dear friends for their valuable suggestion and co-operation for this project work.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | | | | **PAGE NO.** |
|  | **ABSTRACT** | | | | iv |
| **1.** | **INTRODUCTION** | | | | 1 |
|  | 1.1 | LIP MOVEMENT ANALYSIS | | | 1 |
|  | 1.2 | TWO-STREAM NETWORKS | | | 2 |
|  | 1.3 | RANDOM FOREST | | | 3 |
|  | 1.4 | DECISION TREES | | | 3 |
|  | 1.5 | REINFORCEMENT LEARNING | | | 4 |
|  | 1.6 | RECURRENT NEURAL NETWORK | | | 4 |
|  | 1.7 | CNN MODEL | | | 5 |
| **2.** | **LITERATURE SURVEY** | | | | 7 |
|  | 2.1 | INFERENCES | | | 15 |
| **3.** | **DESIGN THINKING** | | | | 16 |
|  | 3.1 | EMPATHY MAP | | | 16 |
|  | 3.2 | CUSTOMER JOURNEY MAP | | | 17 |
| **4.** | **PROPOSED METHODOLOGY** | | | | 19 |
|  | 4.1 | LONG SHORT-TERM MEMORY | | | 19 |
|  | 4.2 | CNN AND GRU MODEL | | | 20 |
|  | 4.3 | ARCHITECTURE DIAGRAM | | | 21 |
|  |  | 4.3.1 | | USE CASE DIAGRAM | 21 |
|  |  | 4.3.2 | | CLASS DIAGRAM | 22 |
|  |  | 4.3.3 | | SEQUENCE DIAGRAM | 23 |
|  |  | 4.3.4 | | PROPOSED DESIGN | 24 |
|  | 4.4 | MODULES | | | 25 |
|  |  | 4.4.1 | MODULE DESCRIPTION | | 25 |
| **5.** | **SYSTEM REQUIREMENTS** | | | | 27 |
|  | 5.1 | SOFTWARE REQUIREMENTS | | | 27 |
|  |  | 5.1.1 | | PYTHON LANGUAGE | 27 |
|  |  | 5.1.2 | | GOOGLE COLABORATORY | 28 |
|  |  | 5.1.3 | | PYTHON PACKAGES | 30 |
| **6.** | **IMPLEMENTATION AND RESULTS** | | | | 32 |
|  | 6.1 | IMPLEMENTATION | | | 32 |
|  |  | 6.1.1 | | DATA PREPROCESSING | 32 |
|  |  | 6.1.2 | | MODEL ARCHITECTURE | 33 |
|  | 6.2 | EXPERIMENTAL RESULTS | | | 34 |
|  |  | 6.2.1 | | DEEP NEURAL NETWORK | 34 |
|  |  | 6.2.2 | | CTC LOSS | 36 |
|  |  | 6.2.3 | | CALLBACK FUNCTION | 36 |
|  |  | 6.2.4 | | MAKING PREDICTION | 38 |
|  |  | 6.2.5 | | TESTING A VIDEO | 39 |
|  |  | 6.2.6 | | ACCURACY | 41 |
| **7.** | **CONCLUSION AND FUTURE SCOPE** | | | | 42 |
|  | 7.1 | CONCLUSION | | | 42 |
|  | 7.2 | FUTURE ENHANCEMENTS | | | 42 |
|  | **APPENDICES** | | | | 44 |
|  | **REFERENCES** | | | | 45 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **FIGURE DESCRIPTION** | **PAGE NO.** |
| 3.1 | EMPATHY MAP | 17 |
| 3.2 | CUSTOMER JOURNEY MAP | 18 |
| 4.1 | USE CASE DIAGRAM | 22 |
| 4.2 | CLASS DIAGRAM | 23 |
| 4.3 | SEQUENCE DIAGRAM | 24 |
| 4.4 | PROPOSED DESIGN | 24 |
| 6.1 | VIDEO LOADING | 33 |
| 6.2 | MODEL ARCHITECTURE | 34 |
| 6.3 | DESIGNING DEEP NEURAL NETWORK | 35 |
| 6.4 | MAKING PREDICTION | 38 |
| 6.5 | TESTING A VIDEO | 40 |

**ABSTRACT**

Speech recognition is unreliable in noisy places, threatens privacy and safety around strangers, and is inaccessible to people with speech impairments. Lip reading can alleviate many of these challenges, but current silent lip reading speech recognition devices are prone to errors. Lip reading is the ability to understand spoken language by observing lip movements. It can be challenging because many sounds look similar on the lips, and not all speech sounds are visible. Additionally, factors like distance, lighting, and facial hair can affect the accuracy of lip reading.Automatic lip reading systems have shown promising results in converting lips into text. This project proposes an improved lip reading model that combines 3D Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) to improve the accuracy and robustness of the reading. This allows the model to analyze successive frames of lip movements and simultaneously capture both spatial and temporal properties. The GRU component is used to learn long term dependencies and capture temporal dynamics in lips.

**CHAPTER 1**

**INTRODUCTION**

In our increasingly noisy and bustling world, clear communication can be a formidable challenge, especially in environments where noise levels drown out spoken words. It's all too common for individuals to struggle with hearing and comprehending words correctly. This difficulty arises because speech recognition technology, which relies on capturing the precise sounds of spoken words, can often misinterpret or miss crucial information in the midst of this auditory chaos.