**Speech-To-Text Using RNN**

*(B. Tech. Project-2)*

*A REPORT*

*Submitted by*

Patel Rudrakumar D.

(21ITUOS011)

*for the partial fulfilment of the requirements for Semester –VII of*

BACHELOR OF TECHNOLOGY (INFORMATION TECHNOLOGY)

*Under the guidance of*

Prof. D. P. Vegda

Dharmsinh Desai University, Nadiad.



Department of Information Technology

Faculty of Technology,

# DHARMSINH DESAI UNIVERSITY

NADIAD 387001

November,2024

**Candidate Disclosure on the Use of AI Tools**

In the process of writing this report, we used the following AI tools and technologies:

In Build AI Features, GitHub copilot,

1. Grammarly (Premium version) was used to generate an outline for this paper and to correct errors in spelling, grammar, and mechanics.

# Candidate’s Declaration

We declare that the dissertation (for B.Tech in Information Technology) titled “Speech-To-Text Using RNN” is our own work being conducted under the guidance and supervision of Prof. D. P. Vegda.

We further declare that to the best of our knowledge; this dissertation does not contain any part of work which has been submitted for the award of any degree either in this University or in any other University without proper citation.

Signature

Rudrakumar. D. Patel

# CERTIFICATE

This is to certify that this Report of B. Tech. Project2 submitted for partial fulfillment of

B. Tech Semester- VII is a record of the work carried out by

1. RUDRAKUMAR DINESHBHAI PATEL

ID No. 21ITUOS011, B. Tech. Sem – VII (Information Technology):2024-25

Guide HoD

Prof. X. Y. Surname Prof. Dr. V. K. Dabhi

Associate/Assistant Professor, Head, Dept. of Information Technology

Department of Information Technology Faculty of Technology

Dharmsinh Desai University, Dharmsinh Desai University

Nadiad–387001, INDIA Nadiad–387001, INDIA



Department of Information Technology

Faculty of Technology

Dharmsinh Desai University

College Road, Nadiad-387001, INDIA

# Acknowledgment

It is indeed a great pleasure to express our thanks and gratitude to all those who helped us during this project. This project has given us a great opportunity to think, implement and interact with various aspects of the Software Development Life Cycle. We would like to acknowledge all the people who have helped us at one stage or another by providing the much-needed support, encouragement, and groundwork to complete our project.

We express a deep sense of gratitude towards our project guide Prof. Deepak Vegda towards his innovative ideas and earnest effort to make our project a success. It is his sincerity that prompted us throughout the project to do hard work using industry-adopted technologies. Our commitment to the application is the sole result of patience, hard work, and dedication being inspired by him.

A blend gratitude, pleasure, and great satisfaction are what we feel to convey our indebtedness to all those who have directly or indirectly contributed towards the completion of the project.

RUDRAKUMRA. D. PATEL

Dharmsinh Desai University, Nadiad

November, 2024

Rudrapatel2992003@gmail.com

-i-

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Abstract

Speech-To-Text Using RNN

Project2 by Rudrakumar. D. Patel

at

Dharmsinh Desai University, November 2024

Speech-to-text technology has become a fundamental component of many applications, ranging from accessibility solutions to automated customer service. This work discusses the effectiveness of using recurrent neural networks for accurate speech-to-text conversion with great efficiency. Audio data in its sequential nature is one of the best inputs toward capturing temporal dependencies, ideal for processing continuous speech streams. This paper deals with training RNNs on a diverse speech dataset containing challenges such as variable-length input sequences, noise, and diverse accents. A many-to-many RNN model configuration was adopted in this work to transcribe spoken language to text with minimal padding to avoid altering the integrity of the data. Our results were successful, and the model proved highly accurate in real-world application scenarios, thereby establishing the feasibility of RNNs as a promising approach for the development of speech-to-text systems. This work therefore shows the strength of models based on RNN in human-computer interaction improvement, as well as their wide applicability in multilingual environments and even noisy environments.

-iii-

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Table of Contents

[Acknowledgment i](#_TOC_250020)

[Abstract ii](#_TOC_250019)

[Table of Contents iii](#_TOC_250018)

[List of Tables iv](#_TOC_250017)

[List of Figures v](#_TOC_250016)

[Abbreviations vi](#_TOC_250015)

1. Introduction to 1
   1. [Introduction to the Research Problem 1](#_TOC_250014)
   2. [Motivation for the Research Work 1](#_TOC_250013)
   3. [Objectives and Scope of the Research Work 1](#_TOC_250012)
2. [Background Theory 3](#_TOC_250011)
   1. [Give appropriate title 3](#_TOC_250010)
   2. [Give appropriate title 3](#_TOC_250009)
3. [Review of Literature. 5](#_TOC_250008)
   1. [Give appropriate title 5](#_TOC_250007)
   2. [Give appropriate title 5](#_TOC_250006)
4. [Analysis and Findings 7](#_TOC_250005)
5. [Proposed Work](#_TOC_250004)
   1. Solution design 8
   2. Implementation Details 8
   3. Experiments and Results 9
6. [Conclusions 10](#_TOC_250003)

[References 11](#_TOC_250002)

[Research Paper 12](#_TOC_250001)

[Curriculum Vitae 13](#_TOC_250000)

-iv-

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# List of Tables

Table 1 Sample Table 7

-v-

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# List of Figures

Figure 1 Sample Image 7

-vi-

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Abbreviations

STT - Speech to Text

RNN - Recurrent Neural Network

ASR - Automatic Speech Recognition

MFCC - Mel-frequency Cepstral Coefficients (used for feature extraction)

WER - Word Error Rate (performance metric)

CTC - Connectionist Temporal Classification (loss function in ASR)

GRU - Gated Recurrent Unit (a type of RNN cell)

LSTM - Long Short-Term Memory (another type of RNN cell)

For a Speech-to-Text (STT) project using Recurrent Neural Networks (RNNs), the goal is to convert audio into accurate textual transcription. RNNs are ideal for this task due to their ability to handle sequential data. Features are often extracted from audio using Mel-frequency Cepstral Coefficients (MFCC) to capture essential sound characteristics. During training, the Connectionist Temporal Classification (CTC) loss function is commonly used to align the input audio with the output transcription. Performance can be measured using Word Error Rate (WER), which indicates accuracy by comparing transcribed text with reference text. Advanced RNN architectures like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) help manage long audio sequences by reducing issues with vanishing gradients. For optimal performance, training is often accelerated using Graphics Processing Units (GPUs).

-vii-

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Introduction

## Introduction to the Research Problem

This project focuses on developing a Speech-to-Text (STT) system for the Gujarati language using Recurrent Neural Networks (RNNs). Speech-to-Text technology is essential for creating voice-enabled applications, assisting with automated transcription, and supporting accessibility tools for diverse language speakers. While there are well-established STT systems for languages like English, creating one for Gujarati presents unique challenges, given the complexity and distinct features of the language.

RNNs are particularly suitable for this project because they can process sequential data, making them effective at capturing the flow and structure of speech. In this system, advanced RNN structures like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are used to handle longer audio sequences and improve transcription accuracy. Connectionist Temporal Classification (CTC) is applied as the loss function to align audio with text, especially since spoken and written forms may have different lengths.

## Motivation for the Research Work

The motivation for developing a Speech-to-Text (STT) system for the Gujarati language stems from the increasing demand for technology that supports regional languages. While popular languages like English, Spanish, and Chinese have robust STT systems, less attention has been given to languages like Gujarati, which has a large number of speakers in India and worldwide. Creating an STT system for Gujarati not only supports inclusivity but also empowers native speakers by enabling them to interact with technology in their own language. This is especially important for individuals who may not be fluent in English or prefer using their mother tongue in digital environments.

Moreover, as voice technology becomes more integrated into daily life—from smart assistants to automated transcription tools—the need for localized language models grows. An STT system for Gujarati could enhance accessibility in fields like education, customer service, and content creation, providing significant benefits to those in Gujarati-speaking communities. This project aims to fill this gap by building a reliable Gujarati STT model using Recurrent Neural Networks (RNNs). By contributing to technology in an underrepresented language, this research not only broadens the reach of digital services but also helps preserve and promote the use of Gujarati in modern applications.

## Objectives and Scope of the Research Work

The primary objective of this research is to develop an efficient Speech-to-Text (STT) system for the Gujarati language using Recurrent Neural Networks (RNNs). This system seeks to accurately convert spoken Gujarati into written text, addressing unique linguistic elements such as phonetics, tone, and syntax that are specific to Gujarati. A key goal is to achieve a low Word Error Rate (WER), ensuring that the transcriptions are both precise and reliable, especially for use in practical applications.

-1-

B.Tech. Dissertation – 2016-17, Department of Information Technology, Dharmsinh Desai University

1. *Introduction*

The scope of this research covers building a model that can handle various accents, speech speeds, and pronunciations within Gujarati, making it adaptable for real-world use. This includes feature extraction from audio, training the model with Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU) for handling long audio sequences, and evaluating the model's accuracy on a diverse dataset. Although this project focuses on the Gujarati language, the techniques and methods applied here could serve as a framework for developing STT systems for other regional languages.

Overall, this research aims to contribute to the technological development of voice applications in Gujarati, providing a foundation for future advancements in accessibility, automation, and digital inclusivity for native speakers.

-2-

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Background Theory

## Give appropriate title

Write here. Use paragraph style.

Write here. Use paragraph style.

## Give appropriate title

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

-3-

M.Tech. Dissertation – 2016-17, Department of Information Technology, Dharmsinh Desai University

1. *Background Theory*

Write here. Use paragraph style.

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Review of Literature

## Give appropriate title

Write here. Use paragraph style.

Write here. Use paragraph style.

## Give appropriate title

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

-5-

M.Tech. Dissertation – 2016-17, Department of Information Technology, Dharmsinh Desai University

1. *Review of Literature*

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Analysis and Findings

**Table 1 Sample Table**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |



**Figure 1 Sample Image**

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

-7-

M.Tech. Dissertation – 2016-17, Department of Information Technology, Dharmsinh Desai University

1. *Analysis and Findings*

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

Write here. Use paragraph style.

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Proposed Work

* 1. Write your methodology / algorithm/model setup / dataset characteristics
  2. Write your implementation /results and discuss the results obtained
  3. Write your experiments and results of each experiment here.

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Conclusions

Write here your conclusion and future work.

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# References

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Research Paper

Write research papers in IEEE format or as per suggested by the guide

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University

# Curriculum Vitae

write your recent cv

B.Tech. Project2 – 2024-25, Department of Information Technology, Dharmsinh Desai University