

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Belgaum, Karanataka-590014



Internship report on:

“Lip to Speech Synthesis”

Submitted in partial fulfilment for the award of degree (18CS185)

Bachelor of Engineering, CSE

Submitted by:

Nayil Ahmed Siddique

2AB19CS037



Conducted at:

Varcons Technologies Pvt Ltd



AHM's

**ANJUMAN INSTITUTE OF TECHNOLOGY AND
MANAGEMENT**

Anjumanabad, Bhatkal – 581320

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2022-23

ANJUMAN INSTITUTE OF TECHNOLOGY AND
MANAGEMENT

ANJUMANABAD, BHATKAL

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(Affiliated to VTU, Belgaum, Recognized by AICTE, New Delhi)



CERTIFICATE

This is to certify that the Internship titled **“Lip to Speech Synthesis”** carried out by **Mr. Nayil Ahmed Siddique**, a bonafide student of Anjuman Institute of Technology, in partial fulfillment for the award of **Bachelor of Engineering**, in **Computer Science** under Visvesvaraya Technological University, Belagavi, during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18CSI85)

Signature of Guide

Signature of HOD

Signature of Principal

External Viva:

Name of the examiner

Signature with date

1) _____

2) _____

DECLARATION

I, **Nayil Ahmed Siddique**, final year student of Computer Science, Anjuman Institute of Technology and Management - 581320, declare that the Internship has been successfully completed, in **Varcons Technologies Pvt Ltd**. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Computer Science, during the academic year 2022-2023.

Date: **23rd September 2022**

Place: **Remote Internship**

USN: **2AB19CS037**

Name: **Nayil Ahmed Siddique**



Varcons Technologies Pvt Ltd

Date: **23rd August, 2022**

Name: **Nayil Ahmed Siddique**

USN: **2AB19CS037**

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python(Research Based)** Internship position with **Varcons Technologies Pvt Ltd**, effective Start Date **23rd August, 2022**, All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python(Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!

Sincerely,

Spoorthi H C

Director

VARCONS TECHNOLOGIES PVT LTD

213, 2st Floor,

18 M G Road, Ulsoor,

Bangalore-560001

ACKNOWLEDGEMENT

This Internship is a result of accumulated guidance, direction and support of several important persons. We take this opportunity to express our gratitude to all who have helped us to complete the Internship.

We express our sincere thanks to our Principal Dr Fazlurahman sir, for providing us ~~ade~~ facilities to undertake this Internship.

We would like to thank our Head of Department, Prof. Anvar Shathik, Computer Science, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank the providers of our Software Services for guiding us during the period of internship.

We express our deep and profound gratitude to our guide, Prof. Prashant Naik, for his keen interest and encouragement at every step in completing the Internship.

We express our gratitude towards our trainer Mr. Prajwal S Madhav, for his continued support in the progression of this internship a till the project completion phase.

We would like to thank all the faculty members of our department for the support extended during the course of Internship.

We would like to thank the non-teaching members of our department, for helping us during the Internship.

Last but not the least, we would like to thank our parents and friends without whose constant help, the completion of Internship would have not been possible.

Nayil Ahmed Siddique

2AB19CS037

CONTENT OF THE PROJECT

CHAPTERS No.	Page
Abstract	i
1. COMPANY PROFILE.....	1
➤ 1.1 A Brief History of Varcons Technologies Pvt Ltd.....	1
2. ABOUT COMPANY	3
➤ 2.1 Products of Varcons Technologies	3
➤ 2.2 Department and services offered	5
3. INTRODUCTION.....	7
➤ 3.1 Introduction to Machine Language	7
➤ 3.2 Classification of Machine Learning	7
➤ 3.3 Categorizing based on output values	9
➤ 3.4 Problem statement	10
4. SYSTEM ANALYSIS.....	11
➤ 4.1 Existing System	11
➤ 4.2 Proposed System	11
➤ 4.3 Objective of the system	11
5. REQUIREMENT ANALYSIS.....	12
➤ 5.1 Hardware Requirement Specification.....	12
➤ 5.2 Hardware Requirement Specification	12
6. DESIGN ANALYSIS	13
➤ 6.1 Web Application.....	13
➤ 6.1.1 Flask Framework.....	13
7. IMPLEMENTATION	14
➤ 7.1 Web Application.....	15
➤ 7.1.1 app.py.....	15
➤ 7.1.2 index.html.....	17

➤ 7.2 Model Training.....	19
➤ 7.3 Model Testing.....	19
8. SNAPSHOTS.....	21
9. CONCLUSION.....	23
REFERENCES.....	24

ABSTRACT

The project developed is concerned with application of Machine Learning. The project is titled as “Lip to Speech synthesis”.

Humans involuntarily tend to infer parts of the conversation from lip movements when the speech is absent or corrupted by external noise. In this work, we explore the task of lip to speech synthesis, i.e., learning to generate natural speech given only the lip movements of a speaker. Acknowledging the importance of contextual and speaker specific cues for accurate lip-reading, we take a different path from existing works. We focus on learning accurate lip sequences to speech mappings for individual speakers in unconstrained, large vocabulary settings. To this end, we collect and release a large-scale benchmark dataset, the first of its kind, specifically to train and evaluate the single speaker lip to speech task in natural settings. We propose a novel approach with key design choices to achieve accurate, natural lip to speech synthesis in such unconstrained scenarios for the first time. Extensive evaluation using quantitative, qualitative metrics and human evaluation shows that our method is four times more intelligible than previous works in this space.

CHAPTER 1

COMPANY PROFILE

1.1 A Brief History of Varcons Technologies Pvt Ltd

Varcons Technologies Pvt Ltd, was incorporated with a goal, *“To provide high quality and optimal Technological Solutions to business requirements of our clients”*. Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions provided to these requirements are different as well. They focus on clients requirements and provide them with tailor made technological solutions. They also understand that Reach of their Product to its targeted market or the automation of the existing process into e-client and simple process are the key features that our clients desire from Technological Solution they are looking for and these are the features that we focus on while designing the solutions for their clients.

Sarvamoola Software Services is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Sarvamoola Software Services specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements.

Varcons Technologies Pvt Ltd, strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the clients demand better. Varcons Technologies Pvt Ltd work with their clients and help them to define their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group comprising of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put it in one sentence ” Technology helps you to Delight your Customers” and that is what we want to achieve.

CHAPTER 2

ABOUT THE COMPANY

Varcons Technologies Pvt Ltd is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Varcons Technologies Pvt Ltd specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements. The organization where they have a right mix of professionals as a stakeholders to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solution. Motto of our organization is to “Collaborate with our clients to provide them with best Technological solution hence creating Good Present and Better Future for our client which will bring a cascading a positive effect in their business shape as well”. Providing a Complete suite of technical solutions is not just our tag line, it is Our Vision for Our Clients and for Us, We strive hard to achieve it.



2.1 Products of Varcons Technologies

Android Apps

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java (and/or Kotlin; or other such option) programming language using the Android software development kit (SDK), but other development environments are also available, some such as Kotlin support the exact same Android APIs (and bytecode), while others such as Go have restricted API access.

The Android software development kit includes a comprehensive set of

development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and tutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows 7 or later. As of March 2015, the SDK is not available on Android itself, but software development is possible by using specialized Android applications.

Web Application

It is a client–server computer program in which the client (including the user interface and client-side logic) runs in a web browser. Common web applications include web mail, online retail sales, online auctions, wikis, instant messaging services and many other functions. web applications use web documents written in a standard format such as HTML and JavaScript, which are supported by a variety of web browsers. Web applications can be considered as a specific variant of client–server software where the client software is downloaded to the client machine when visiting the relevant web page, using standard procedures such as HTTP. The Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, and acts as the universal client for any web application. The use of web application frameworks can often reduce the number of errors in a program, both by making the code simpler, and by allowing one team to concentrate on the framework while another focuses on a specified usecase. In applications which are exposed to constant hacking attempts on the Internet, security-related problems can be caused by errors in the program.

Frameworks can also promote the use of best practices such as GET after POST. There are some who view a web application as a two-tier architecture. This can be a “smart” client that performs all the work and queries a “dumb” server, or a “dumb” client that relies on a “smart” server. The client would handle the presentation tier, the server would have the database (storage tier), and the business logic (application tier) would be on one of them or on both. While this increases the scalability of the applications and separates the display and the database, it still doesn’t allow for true specialization of layers, so most applications will outgrow this model. An emerging strategy for application

software companies is to provide web access to software previously distributed as local applications. Depending on the type of application, it may require the development of an entirely different browser-based interface, or merely adapting an existing application to use different presentation technology. These programs allow the user to pay a monthly or yearly fee for use of a software application without having to install it on a local hard drive. A company which follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry.

Security breaches on these kinds of applications are a major concern because it can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and there are some key operational areas that must be included in the development process. This includes processes for authentication, authorization, asset handling, input, and logging and auditing. Building security into the applications from the beginning can be more effective and less disruptive in the long run.

Web design

It encompasses many different skills and disciplines in the production and maintenance of websites. The different areas of web design include web graphic design; interface design; authoring, including standardized code and proprietary software; user experience design; and

search engine optimization. The term web design is normally used to describe the design process relating to the front-end (client side) design of a website including writing mark up. Web design partially overlaps web engineering in the broader scope of web development. Web designers are expected to have an awareness of usability and if their role involves creating markup then they are also expected to be up to date with web accessibility guidelines. Web design partially overlaps web engineering in the broader scope of web development.

2.3 Department and services offered

Varcons Technologies Pvt Ltd plays an essential role as an institute, the level of education, development of student's skills are based on their trainers. If you do

not have a good mentor then you may lag in many things from others and that is why we at Varcons Technologies Pvt Ltd gives you the facility of skilled employees so that you do not feel unsecured about the academics. Personality development and academic status are some of those things which lie on mentor's hands. If you are trained well then you can do well in your future and knowing its importance of Varcons Technologies Pvt Ltd always tries to give you the best.

They have a great team of skilled mentors who are always ready to direct their trainees in the best possible way they can and to ensure the skills of mentors we held many skill development programs as well so that each and every mentor can develop their own skills with the demands of the companies so that they can prepare a complete packaged trainee.

Services provided by Varcons Technologies Pvt Ltd

- Core Java and Advanced Java
- Web services and development
- Dot Net Framework
- Python
- Selenium Testing
- Conference / Event Management Service
- Academic Project Guidance
- On The Job Training
- Software Train

CHAPTER 3

INTRODUCTION

3.1 Introduction to Machine Language

Arthur Samuel, an early American leader in the field of computer gaming and artificial intelligence, coined the term “Machine Learning” in 1959 while at IBM. He defined machine learning as “the field of study that gives computers the ability to learn without being explicitly programmed “. However, there is no universally accepted definition for machine learning. Different authors define the term differently. We give below two more definitions.

Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data.

The field of study known as machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.

Definition: A computer program which learns from experience is called a machine learning program or simply a learning program.

3.2 Classification of Machine Learning

Machine learning implementations are classified into four major categories, depending on the nature of the learning “signal” or “response” available to a learning system which are as follows:

1. Supervised learning:

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. The given data is labeled. Both classification and regression problems are supervised learning problems.

Example: Consider the following data regarding patients entering a clinic . The data consists of the gender and age of the patients and each patient is labeled as “healthy” or “sick”.

gender	age	label
M	48	sick
M	67	sick
F	53	healthy
M	49	sick
F	32	healthy
M	34	healthy
M	21	healthy

2. Unsupervised learning

Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses. In unsupervised learning algorithms, classification or categorization is not included in the observations.

Example: Consider the following data regarding patients entering a clinic. The data consists of the gender and age of the patients.

gender	age
M	48
M	67
F	53
M	49
F	34
M	21

3. Reinforcement learning

Reinforcement learning is the problem of getting an agent to act in the world so as to maximize its rewards.

A learner is not told what actions to take as in most forms of machine learning but instead must discover which actions yield the most reward by trying them.

For example: Consider teaching a dog a new trick: we cannot tell it what to do, but we can reward/punish it if it does the right/wrong thing.

4. Semi-supervised learning

Where an incomplete training signal is given: a training set with some (often many) of the target outputs missing. There is a special case of this principle known as Transduction where the entire set of problem instances is known at learning time, except that part of the targets are missing. Semi-supervised learning is an approach to machine learning that combines small labeled data with a large amount of unlabeled data during training. Semi-supervised learning falls between unsupervised learning and supervised learning.

3.3 Categorizing based on required Output

Another categorization of machine learning tasks arises when one considers the desired output of a machine-learned system:

Classification: When inputs are divided into two or more classes, the learner must produce a model that assigns unseen inputs to one or more (multi-label classification) of these classes. This is typically tackled in a supervised way. Spam filtering is an example of classification, where the inputs are email (or other) messages and the classes are “spam” and “not spam”.

Regression: Which is also a supervised problem, A case when the outputs are continuous rather than discrete.

Clustering: When a set of inputs is to be divided into groups. Unlike in classification, the groups are not known beforehand, making this typically an unsupervised task.

Machine Learning comes into the picture when problems cannot be solved using typical approaches. ML algorithms combined with new computing technologies promote scalability and improve efficiency.

3.4 Problem statement:

A source code of a Lip to Speech synthesis- is provided, the goal is to test all the features, determine the overall accuracy of the features and try to improve the model. The current accuracy rate: 84%.

CHAPTER 4

SYSTEM ANALYSIS

4.1. Existing System

Currently, upon experiencing any symptoms the patient searches it on the search engine like Google, Bing etc. which deteriorates the health even more due to the not so accurate prediction of the disease.

The existing system is quite sparse and highly unstable as it does not provide the concerned patient with a definitive answer which eventually leads to stress and tension affecting the health of the user.

Due to presence of many sources, the results are multiple and contradictory on internet which makes the

4.2 Proposed System

The proposed system is a chat bot which is trained with many different sets of symptoms and the diseases associated with it. It is highly accurate in predicting the diseases and learning new symptoms at the same time.

Datasets can be provided to increase the accuracy of the prediction in very less time and can be trained within less time.

4.3 Objective of the System

The objective of the proposed system is to accurately predict the disease through a chat bot which will learn new patterns of symptoms at the same time and train itself for further conversations.

Improving accuracy of the prediction by training the model with large datasets and successfully predict the diseases and suggest the required medicines.

CHAPTER 5

REQUIREMENT ANALYSIS

5.1 Hardware Requirement Specification

Dual Core CPU / Quad Core CPU

8 GB RAM / 16 GB RAM

20 GB / 50 GB Free Disk Space

NVIDIA® GPU card with CUDA® architectures 3.5 or higher

5.2 Software Requirement Specification

Operating Systems:

- Ubuntu 16.04 or higher (64-bit)
- macOS 10.12.6 (Sierra) or higher (64-bit) (no GPU support)
- Windows Native - Windows 7 or higher (64-bit)
- Windows WSL2 - Windows 10 19044 or higher (64-bit)

Database requirements:

- Oracle 19
- PostgreSQL 12.8 or newer

Modules requirements:

- Python 3
- Flask

CHAPTER 6

DESIGN ANALYSIS

Modules Description

The proposed system is going to implemented using two specific modules and each module will be divided into sub modules according to requirement.

6.1 Web Application

The web application is designed using python based Flask framework which act as a frontend of the project. It renders a single page named 'index.html' which contains the UI of the chat bot.

6.1.1 Flask Framework

Flask is an API of Python that allows us to build up web-applications. It was developed by Armin Ronacher. Flask's framework is more explicit than Django's framework and is also easier to learn because it has less base code to implement a simple web-Application. A Web-Application Framework or Web Framework is the collection of modules and libraries that helps the developer to write applications without writing the low-level codes such as protocols, thread management, etc. Flask is based on WSGI(Web Server Gateway Interface) toolkit and Jinja2 template engine

CHAPTER 7

IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and its constraints on implementation, design of methods to achieve the changeover and an evaluation of changeover methods as part from planning.

Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

7.1 Web Application

7.1.1 app.py

```
from flask import Flask, render_template, request, jsonify,
send_from_directory, send_file, flash, redirect
from werkzeug.utils import secure_filename
import os
from model import *

UPLOAD_FOLDER = './uploads'
ALLOWED_EXTENSIONS = {'txt', 'pdf', 'png', 'jpg', 'jpeg', 'gif', 'mp4', 'mp3',
'wav', 'ogg'}

app = Flask(__name__)
app.config['UPLOAD_FOLDER'] = UPLOAD_FOLDER

# @app.route('/upload')
# def returnHTML():
#     return render_template('index.html')

# @app.route('/uploader', methods = ['GET', 'POST'])
# def upload_file():
#     if request.method == 'POST':
#         f = request.files['file']
#         f.save(secure_filename(f.filename))
#         return 'file uploaded successfully'

def allowed_file(filename):
    return '.' in filename and \
        filename.rsplit('.', 1)[1].lower() in ALLOWED_EXTENSIONS

@app.route('/', methods=['GET', 'POST'])
def upload_file():
    if request.method == 'POST':
        # check if the post request has the file part
        if 'file' not in request.files:
            flash('No file part')
            return redirect(request.url)
        file = request.files['file']
        # If the user does not select a file, the browser submits an
        # empty file without a filename.
        if file.filename == ":
```

```

        flash('No selected file')
        return redirect(request.url)
    if file and allowed_file(file.filename):
        filename = secure_filename(file.filename)
        file.save(os.path.join(app.config['UPLOAD_FOLDER'], filename))
        transcription = readLip(filename)
        full_path = app.root_path
        print(full_path)
        return render_template("fileUpload.html", transcription = transcription)
        # return send_from_directory(path=full_path,
filename="./__temp__.mp4")
        # return send_file(full_path+"/__temp__.mp4", as_attachment=True)
        # return readLip(filename)
    return render_template("./fileUpload.html")

@app.route('/download', methods=['GET', 'POST'])
def download():
    full_path = app.root_path
    # return send_from_directory(full_path, filename)
    filename="__temp__.mp4"
    return send_file(filename, as_attachment=True)

@app.route('/api', methods = ['GET'])
def returnAscii():
    d = { }
    return { 'output': ord(str(request.args['query'])) }

if __name__ == "__main__":
    app.secret_key = 'super secret key'
    app.config['SESSION_TYPE'] = 'filesystem'

    app.run(debug = True)

```


7.1.2 Index.html

```
<!DOCTYPE html>
<html>

<head>
  <link rel="stylesheet" type="text/css" href="/static/index.css" />
  <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.2.1/jquery.min.js"></script>
</head>

<body>
  <h1>Healthcare Chatbot Build using rasa</h1>
  <div>
    <div id="chatbox">
      <p class="botText">
        <span>Please try typing full sentences as I am still learning!</span>
      </p>
      <p class="botText">
        <span>I am a your Health assistant.</span>
      </p>
      <p class="botText">
        <span>Hi There! how can i help you. What Symptoms you are facing</span>
      </p>
    </div>
    <div id="userInput">
      <input id="textInput" type="text" name="msg" placeholder="Message" />
      <input id="buttonInput" type="submit" value="Send" />
    </div>
    <script>
      function getBotResponse() {
        var rawText = $("#textInput").val();
        var userHtml = '<p class="userText"><span> ' + "Me: " + rawText + "</span></p>";
        $("#textInput").val("");
        $("#chatbox").append(userHtml);
        document
          .getElementById("userInput")
          .scrollIntoView({ block: "start", behavior: "smooth" });
        $.get("/get", { msg: rawText }).done(function (data) {
          var botHtml = '<p class="botText"><span> ' + "HealthBot: " + data + "</span></p>";
          $("#chatbox").append(botHtml);
          document.getElementById("userInput").scrollIntoView({ block: "start", behavior:
"smooth" });
        });
      }
      $("#textInput").keypress(function (e) {if (e.which == 13) {getBotResponse();}});
      $("#buttonInput").click(function () {getBotResponse();});
    </script>
  </div>
</body>

</html>
```

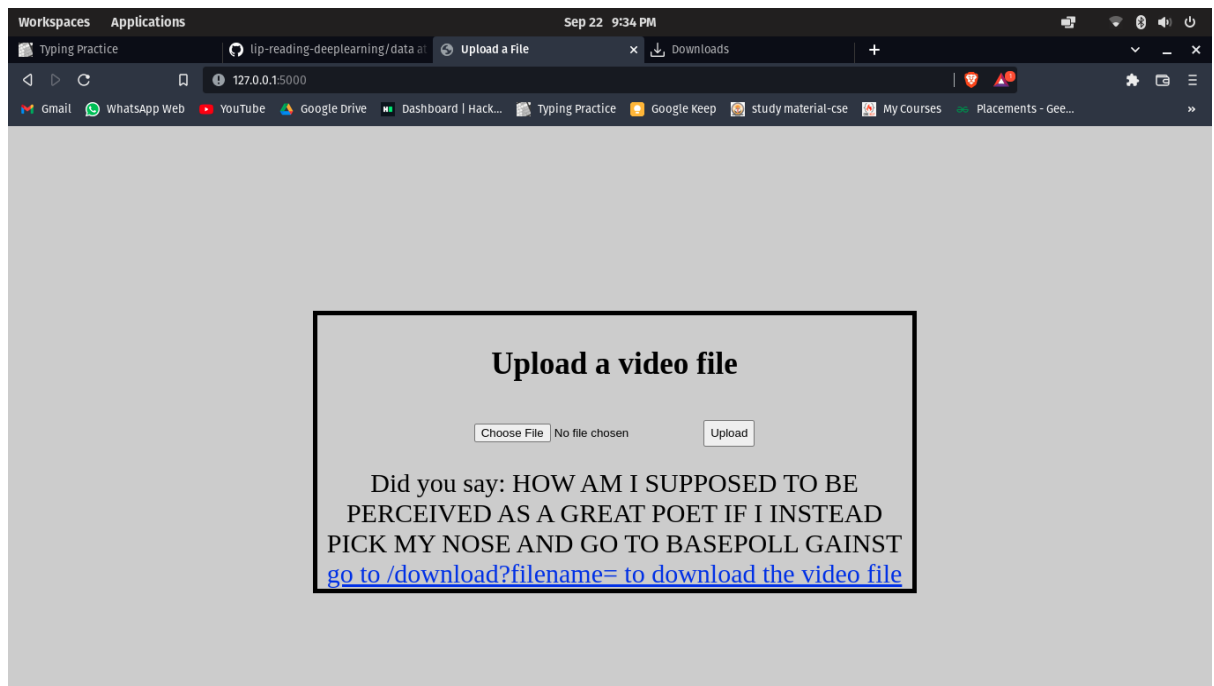


Figure 7.1

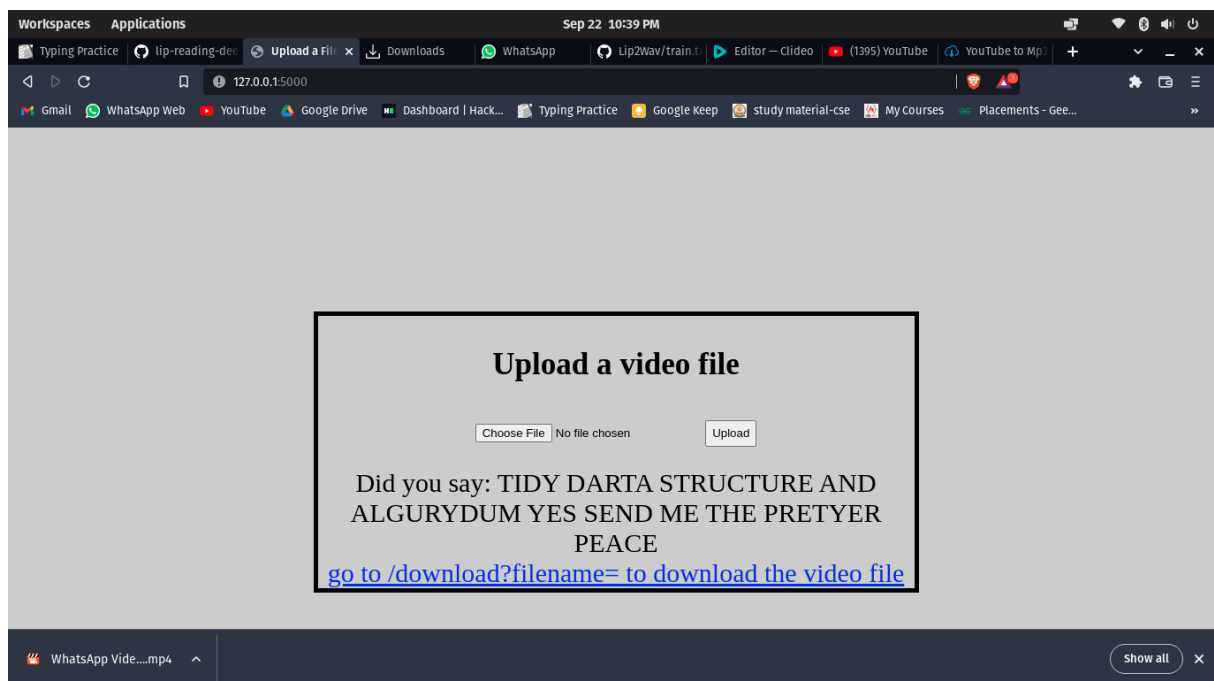
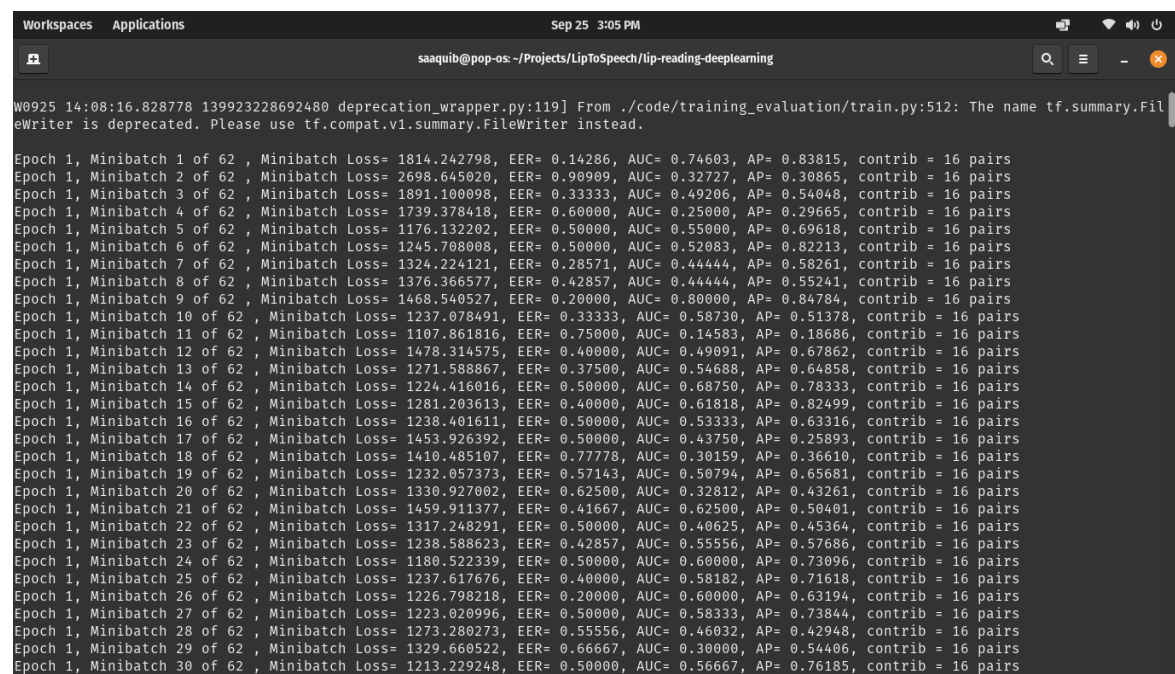


Figure 7.1

7.2 Model Training

The goal of NLU (Natural Language Understanding) is to extract structured information from user messages. This usually includes the user's intent and any entities their message contains. You can add extra information such as regular expressions and lookup tables to your training data to help the model identify intents and entities correctly.

NLU training data consists of example user utterances categorized by intent. To make it easier to use your intents, give them names that relate to what the user wants to accomplish with that intent, keep them in lowercase, and avoid spaces and special characters. Command: `rasa train`



```
W0925 14:08:16.828778 139923228692480 deprecation_wrapper.py:119] From ./code/training_evaluation/train.py:512: The name tf.summary.FileWriter is deprecated. Please use tf.compat.v1.summary.FileWriter instead.
Epoch 1, Minibatch 1 of 62, Minibatch Loss= 1814.242798, EER= 0.14286, AUC= 0.74603, AP= 0.83815, contrib = 16 pairs
Epoch 1, Minibatch 2 of 62, Minibatch Loss= 2698.645020, EER= 0.90909, AUC= 0.32727, AP= 0.30865, contrib = 16 pairs
Epoch 1, Minibatch 3 of 62, Minibatch Loss= 1891.100098, EER= 0.33333, AUC= 0.49206, AP= 0.54048, contrib = 16 pairs
Epoch 1, Minibatch 4 of 62, Minibatch Loss= 1739.378418, EER= 0.60000, AUC= 0.25000, AP= 0.29665, contrib = 16 pairs
Epoch 1, Minibatch 5 of 62, Minibatch Loss= 1176.132202, EER= 0.50000, AUC= 0.55000, AP= 0.69618, contrib = 16 pairs
Epoch 1, Minibatch 6 of 62, Minibatch Loss= 1245.708008, EER= 0.50000, AUC= 0.52083, AP= 0.82213, contrib = 16 pairs
Epoch 1, Minibatch 7 of 62, Minibatch Loss= 1324.224121, EER= 0.28571, AUC= 0.44444, AP= 0.50261, contrib = 16 pairs
Epoch 1, Minibatch 8 of 62, Minibatch Loss= 1376.366577, EER= 0.42857, AUC= 0.44444, AP= 0.55241, contrib = 16 pairs
Epoch 1, Minibatch 9 of 62, Minibatch Loss= 1468.540527, EER= 0.20000, AUC= 0.80000, AP= 0.84784, contrib = 16 pairs
Epoch 1, Minibatch 10 of 62, Minibatch Loss= 1237.078491, EER= 0.33333, AUC= 0.58730, AP= 0.51378, contrib = 16 pairs
Epoch 1, Minibatch 11 of 62, Minibatch Loss= 1107.861816, EER= 0.75000, AUC= 0.14583, AP= 0.18686, contrib = 16 pairs
Epoch 1, Minibatch 12 of 62, Minibatch Loss= 1478.314575, EER= 0.40000, AUC= 0.49091, AP= 0.67862, contrib = 16 pairs
Epoch 1, Minibatch 13 of 62, Minibatch Loss= 1271.588867, EER= 0.37500, AUC= 0.54688, AP= 0.64858, contrib = 16 pairs
Epoch 1, Minibatch 14 of 62, Minibatch Loss= 1224.416016, EER= 0.50000, AUC= 0.68750, AP= 0.78333, contrib = 16 pairs
Epoch 1, Minibatch 15 of 62, Minibatch Loss= 1281.203613, EER= 0.40000, AUC= 0.61818, AP= 0.82499, contrib = 16 pairs
Epoch 1, Minibatch 16 of 62, Minibatch Loss= 1238.401611, EER= 0.50000, AUC= 0.53333, AP= 0.63316, contrib = 16 pairs
Epoch 1, Minibatch 17 of 62, Minibatch Loss= 1453.926392, EER= 0.50000, AUC= 0.43750, AP= 0.25893, contrib = 16 pairs
Epoch 1, Minibatch 18 of 62, Minibatch Loss= 1410.485107, EER= 0.77778, AUC= 0.30159, AP= 0.36610, contrib = 16 pairs
Epoch 1, Minibatch 19 of 62, Minibatch Loss= 1232.057373, EER= 0.57143, AUC= 0.50794, AP= 0.65681, contrib = 16 pairs
Epoch 1, Minibatch 20 of 62, Minibatch Loss= 1330.927002, EER= 0.62500, AUC= 0.32812, AP= 0.43261, contrib = 16 pairs
Epoch 1, Minibatch 21 of 62, Minibatch Loss= 1459.911377, EER= 0.41667, AUC= 0.62500, AP= 0.50401, contrib = 16 pairs
Epoch 1, Minibatch 22 of 62, Minibatch Loss= 1317.248291, EER= 0.50000, AUC= 0.40625, AP= 0.45364, contrib = 16 pairs
Epoch 1, Minibatch 23 of 62, Minibatch Loss= 1238.588623, EER= 0.42857, AUC= 0.55556, AP= 0.57686, contrib = 16 pairs
Epoch 1, Minibatch 24 of 62, Minibatch Loss= 1180.522339, EER= 0.50000, AUC= 0.60000, AP= 0.73096, contrib = 16 pairs
Epoch 1, Minibatch 25 of 62, Minibatch Loss= 1237.617676, EER= 0.40000, AUC= 0.58182, AP= 0.71618, contrib = 16 pairs
Epoch 1, Minibatch 26 of 62, Minibatch Loss= 1226.798218, EER= 0.20000, AUC= 0.60000, AP= 0.63194, contrib = 16 pairs
Epoch 1, Minibatch 27 of 62, Minibatch Loss= 1223.020996, EER= 0.50000, AUC= 0.58333, AP= 0.73844, contrib = 16 pairs
Epoch 1, Minibatch 28 of 62, Minibatch Loss= 1273.280273, EER= 0.55556, AUC= 0.46032, AP= 0.42948, contrib = 16 pairs
Epoch 1, Minibatch 29 of 62, Minibatch Loss= 1239.660522, EER= 0.66667, AUC= 0.30000, AP= 0.54406, contrib = 16 pairs
Epoch 1, Minibatch 30 of 62, Minibatch Loss= 1213.229248, EER= 0.50000, AUC= 0.56667, AP= 0.76185, contrib = 16 pairs
```

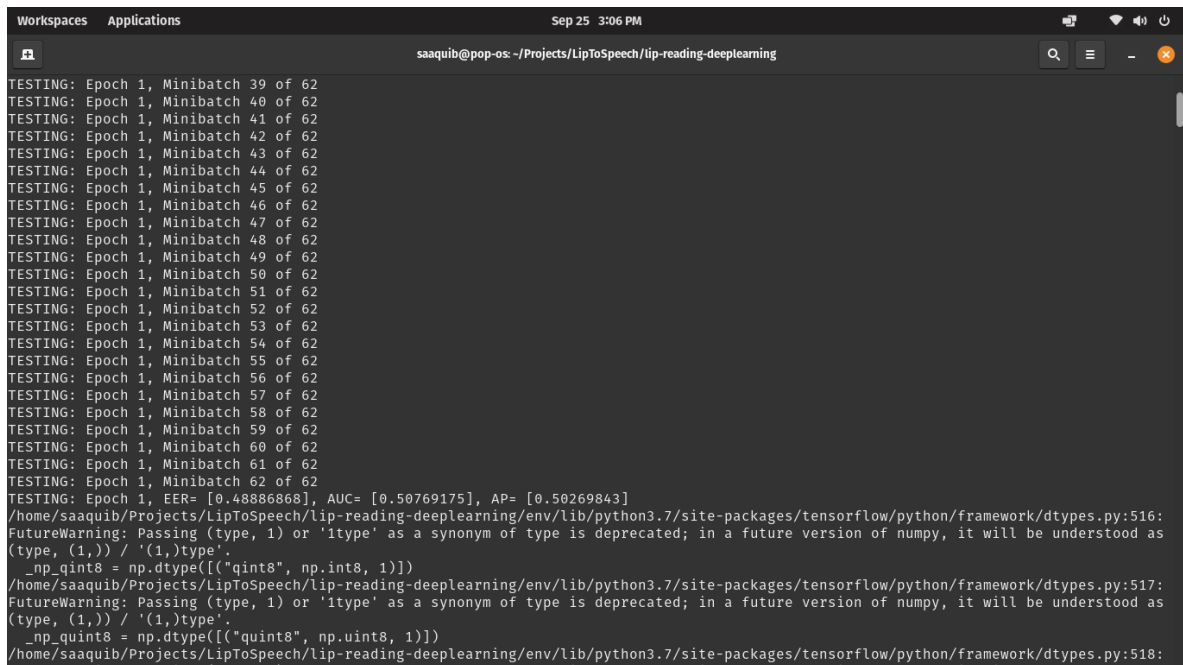
Figure 7.3

7.3 Model Testing

The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately.

2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results



The screenshot shows a terminal window with the title bar 'Workspaces Applications' and 'Sep 25 3:06 PM'. The terminal output displays a series of 'TESTING: Epoch 1, Minibatch' messages, ranging from 39 of 62 to 62 of 62. Following these, it shows performance metrics: 'EER= [0.48886868], AUC= [0.50769175], AP= [0.50269843]'. There are also several 'FutureWarning' messages from numpy regarding deprecated type synonyms. The terminal window has a dark background and standard window controls.

```
TESTING: Epoch 1, Minibatch 39 of 62
TESTING: Epoch 1, Minibatch 40 of 62
TESTING: Epoch 1, Minibatch 41 of 62
TESTING: Epoch 1, Minibatch 42 of 62
TESTING: Epoch 1, Minibatch 43 of 62
TESTING: Epoch 1, Minibatch 44 of 62
TESTING: Epoch 1, Minibatch 45 of 62
TESTING: Epoch 1, Minibatch 46 of 62
TESTING: Epoch 1, Minibatch 47 of 62
TESTING: Epoch 1, Minibatch 48 of 62
TESTING: Epoch 1, Minibatch 49 of 62
TESTING: Epoch 1, Minibatch 50 of 62
TESTING: Epoch 1, Minibatch 51 of 62
TESTING: Epoch 1, Minibatch 52 of 62
TESTING: Epoch 1, Minibatch 53 of 62
TESTING: Epoch 1, Minibatch 54 of 62
TESTING: Epoch 1, Minibatch 55 of 62
TESTING: Epoch 1, Minibatch 56 of 62
TESTING: Epoch 1, Minibatch 57 of 62
TESTING: Epoch 1, Minibatch 58 of 62
TESTING: Epoch 1, Minibatch 59 of 62
TESTING: Epoch 1, Minibatch 60 of 62
TESTING: Epoch 1, Minibatch 61 of 62
TESTING: Epoch 1, Minibatch 62 of 62
TESTING: Epoch 1, EER= [0.48886868], AUC= [0.50769175], AP= [0.50269843]
/home/saaquib/Projects/LipToSpeech/lip-reading-deeplearning/env/lib/python3.7/site-packages/tensorflow/python/framework/dtypes.py:516:
FutureWarning: Passing (type, 1) or 'ltype' as a synonym of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  _np_qint8 = np.dtype [("qint8", np.int8, 1)])
/home/saaquib/Projects/LipToSpeech/lip-reading-deeplearning/env/lib/python3.7/site-packages/tensorflow/python/framework/dtypes.py:517:
FutureWarning: Passing (type, 1) or 'ltype' as a synonym of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  _np_qint8 = np.dtype [("qint8", np.int8, 1)])
/home/saaquib/Projects/LipToSpeech/lip-reading-deeplearning/env/lib/python3.7/site-packages/tensorflow/python/framework/dtypes.py:518:
FutureWarning: Passing (type, 1) or 'ltype' as a synonym of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  _np_qint8 = np.dtype [("qint8", np.int8, 1)])
```

when run as a whole.

Figure 7.4

CHAPTER 8

SNAPSHOTS

8.1 Lip to Speech Snapshots

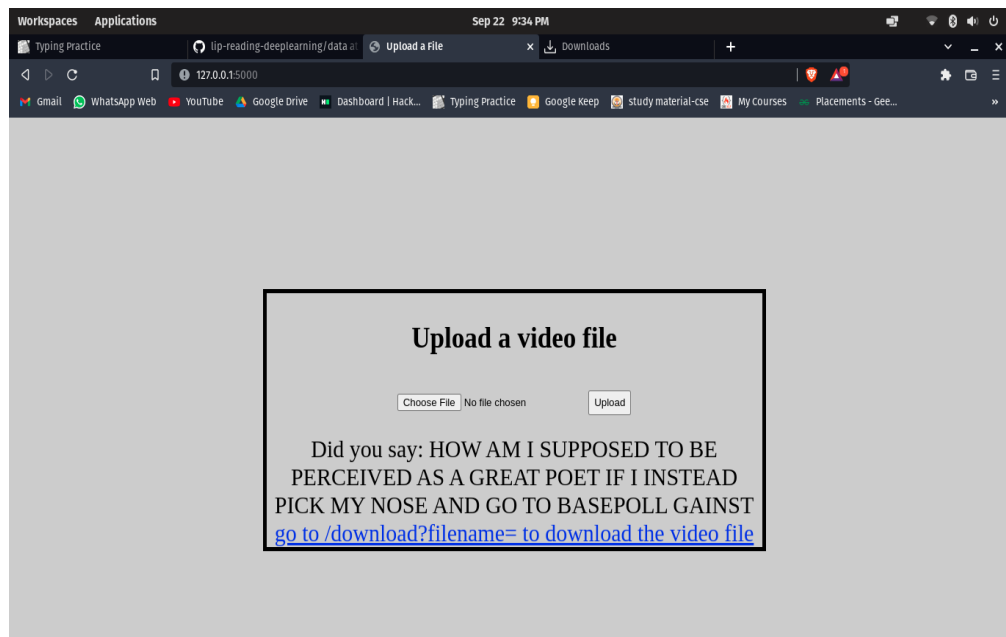


Figure 8.1

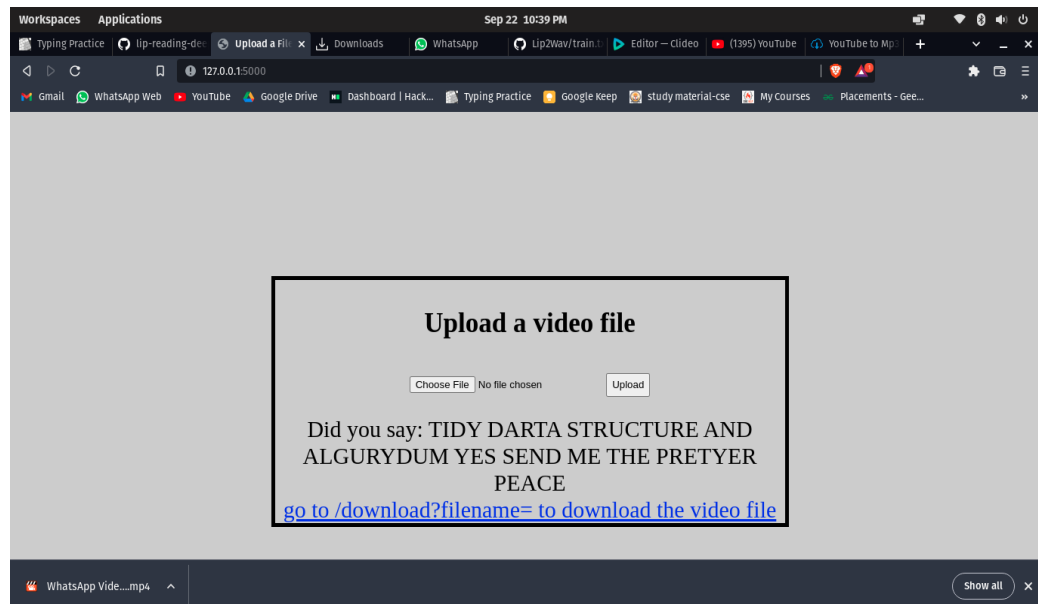


Figure 8.2

```
Workspaces Applications Sep 25 3:03 PM
saaquib@pop-os: ~/Projects/LipToSpeech/lip-reading-deeplearning

ble-libass --enable-libbluray --enable-libbs2b --enable-libcaca --enable-libcdio --enable-libcodecs2 --enable-libdavid --enable-libflite
--enable-libfontconfig --enable-libfreetype --enable-libfribidi --enable-libgme --enable-libgsm --enable-libjack --enable-libmp3lame -
--enable-libmysofa --enable-libopenjpeg --enable-libopenmpt --enable-libopus --enable-libpulse --enable-librabbitmq --enable-librubberba
nd --enable-libshine --enable-lisnappy --enable-libsoxr --enable-lispeex --enable-lisrt --enable-libssh --enable-libtheora --enable-
libtwolame --enable-libvidstab --enable-libvorbis --enable-libvpx --enable-libwebp --enable-libx265 --enable-libxml2 --enable-libxvid -
--enable-libzimg --enable-libzmq --enable-libzvbi --enable-lv2 --enable-omx --enable-openal --enable-openc1 --enable-opengl --enable-sdl
2 --enable-pocketsphinx --enable-libsvg --enable-libmfx --enable-libdc1394 --enable-libdrm --enable-libiec61883 --enable-chromaprint -
--enable-frei0r --enable-libx264 --enable-shared --enable-version3 --disable-doc --disable-programs --enable-libaribb24 --enable-libopen
core_amrnb --enable-libopencore_amrwb --enable-libtesseract --enable-libvo_amrwbenc --enable-lismbclient
libavutil 56. 70.100 / 56. 70.100
libavcodec 58.134.100 / 58.134.100
libavformat 58. 76.100 / 58. 76.100
libavdevice 58. 13.100 / 58. 13.100
libavfilter 7.110.100 / 7.110.100
libswscale 5. 9.100 / 5. 9.100
libswresample 3. 9.100 / 3. 9.100
libpostproc 55. 9.100 / 55. 9.100
[image2 @ 0x556b4593ca80] Pattern type 'glob_sequence' is deprecated: use pattern_type 'glob' instead
Input #0, image2, from './results/mouth/frame_*.png':
  Duration: 00:00:05.12, start: 0.000000, bitrate: N/A
  Stream #0:0: Video: png, gray(pc), 107x99, 25 fps, 25 tbr, 25 tbc
File 'results/mouth.gif' already exists. Overwrite? [y/N] y
Stream mapping:
  Stream #0:0 -> #0:0 (png (native) -> gif (native))
Press [q] to stop, [?] for help
Output #0, gif, to 'results/mouth.gif':
  Metadata:
    encoder      : Lavf58.76.100
  Stream #0:0: Video: gif, gray(pc, progressive), 107x99, q=2-31, 200 kb/s, 25 fps, 100 tbn
  Metadata:
    encoder      : Lavc58.134.100 gif
frame= 128 fps=0.0 q=-0.0 Lsize= 1165kB time=00:00:05.09 bitrate=1875.3kbits/s speed=79.7x
video:1165kB audio:0kB subtitle:0kB other streams:0kB global headers:0kB muxing overhead: 0.001676%
(env) saaquib@pop-os:~/Projects/LipToSpeech/lip-reading-deeplearning$
```

Figure 8.3

CHAPTER 9

CONCLUSION

In this project, we investigated the problem of synthesizing speech based on lip movements. We specifically solved the problem by focusing on individual speakers. We did this in a data-driven learning approach by creating a large-scale benchmark dataset for unconstrained, large vocabulary single-speaker lip to speech synthesis. We formulate the task at hand as a sequence-to-sequence problem, and show that by doing so, we achieve significantly more accurate and natural speech than previous methods. We evaluate our model with extensive quantitative metrics and human studies. All the code and data for our work has been made publicly available . Our work opens up several new directions. One of them would be to examine related works in this space such as lip to text generation from a speaker specific perspective. Similarly, explicitly addressing the dominant issue of homophones can yield more accurate speech. Generalizing to vocabulary outside the typical domain of the speaker can be another fruitful venture. We believe that exploring some of the above problems in a data driven fashion could lead to further useful insights in this space.

REFERENCES

1. Flask Documentation:
<https://flask.palletsprojects.com/en/2.2.x/>
2. Python Documentation:
<https://www.python.org/doc/>
3. Flask Tutorials:
<https://www.youtube.com/c/Coreyms>