VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"Lip to Speech Synthesis"

Submitted in partial fulfilment for the award of degree(18CSI85)

BACHELOR OF ENGINEERING IN
ELECTRONICS AND COMMUNICATION

Submitted by:

SINDHU.S

1SG19EC095

LOGO Ex:

Conducted at
VARCONS TECHNOLOGIES PVT LTD

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SAPTHAGIRI COLLEGE OF ENGINEERING

14/5, Chikkasandra, Hesaraghatta Main Road, Bengaluru– 560057

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SAPTHAGIRI COLLEGE OF ENGINEERING

14/5, Chikkasandra, Hesaraghatta Main Road, Bengaluru– 560057

CERTIFICATE

This is to certify that the Internship titled **“Lip to Speech Synthesis”** carried out by **Mrs. SINDHU S,** a bonafide student of sapthagiri college of engineering, in partial fulfillment for the award of **Bachelor of Engineering**, in **ELECTRONICS AND COMMUNICATION** 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, **SINDHU S**, final year student of Electronics and communication , Sapthagiri college of engineering- 560 082, declare that the Internship has been successfully completed, in **VARCON TECHNOLOGIES PVT LTD**. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Electronis and communication, during the academic year 2022-2023.

Date : 6/10/2023 :

Place : Banglore

USN : 1SG19EC095

NAME : SINDHU S

OFFER LETTER

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, for providing us adequate facilities to undertake this Internship.

We would like to thank our Head of Dept – branch code, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank our (Lab assistant name) Software Services for guiding us during the period of internship.

We express our deep and profound gratitude to our guide, Guide name, Assistant/Associate Prof, for her keen interest and encouragement at every step in completing the Internship.

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 dept, forhelping 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.

SINDHU S

1SG19EC095

ABSTRACT

When we speak, the prosody and content of the speech can be inferred from the movement of our lips. In this work, we explore the task of lip to speech synthesis, i.e., learning to generate speech given only the lip movements of a speaker where we focus on learning accurate lip to speech mappings for multiple speakers in unconstrained, large vo- cabulary settings. We capture the speaker’s voice identity through their facial characteristics, i.e., age, gender, ethnicity and condition them along with the lip movements to generate speaker identity aware speech. To this end, we present a novel method ”Lip2Speech”, with key design choices to achieve accurate lip to speech synthesis in unconstrained scenarios. We also perform various experiments and extensive evaluation using quanti- tative, qualitative metrics and human evaluation.

Table of Contents

CHAPTER 1

COMPANY PROFILE

1. COMPANY PROFILE

A Brief History of VARCONS Technologies

varcons Technologies, 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, 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 work with their clients and help them to defiine 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 itin one sentence ” Technology helps you to Delight your Customers” and that is what we want to achieve.

CHAPTER 2

ABOUT THE COMPANY

2. ABOUT THE COMPANY

Varcons Technologies 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 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.

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 zutorials. 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 specifific 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 use case. 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 is 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 mark up 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.

Departments and services offered

Varcons Technologies 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 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 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.

- 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 Training

CHAPTER 3

INTRODUCTION

3. INTRODUCTION

Introduction to ML

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.

Problem Statement

Given an input video of a speaking face, our objective is to synthesize the speech of the speaking face with the identity of the speaker estimated through their face attributes, i.e., gender, age, ethnicity. The video contains a sequence of image frames, F = (F_1, F_2, F_3, , F_N), and while the synthesized speech con-tains speech frames of the melspectogram, S = (S_1, S_2, S_3, ..., S_T) . And this can be formulated as a sequence-to-sequence learning task[35], where the cor- respondences between the image frames and speech frames are learnt. correspondences capture both the speaker identity and speech content. And, af- ter learning the correspondences, speech frames can be synthesized for the image frames in an auto-regressive manner. It is to be noted that the number of image frames, N and number of speech frames, T does not need to be equal. Con- cretely, each output speech frame, S_t is modelled as a conditional distribution of the previous speech frame and image frames

Vercent Technologies Pvt Ltd

CHAPTER 4

SYSTEM ANALYSIS

4. SYSTEM ANALYSIS

1. **Existing System**

2. **Proposed System**

3. **Objective of the System**

CHAPTER 5

REQUIREMENT ANALYSIS

5. REQUIREMENT ANALYSIS

Hardware Requirement Specification

- TPUs from Google
- Nvidia’s GPUs
- Intel’s CPU, FPGA, ASICs

In this paper, we will discuss the major hardware provided by these giants.

Google Tensor Processing Unit (TPU)

Google’s Tensor Processing Unit (TPU) is an ASIC created by Google to process Neural Networks. Google announced its first hardware in 2016, the TPUs have been in use in its data centers for over a year. Google designed the hardware to work with their open-source software Tensor-flow, an application specifically built for working with Neural Networks. Because of Time-To-Market constraints and the need to work with existing deployments, it was packaged as an external accelerator card that fits into the SATA hard disk slot as a drop-in installation. One motivation for creating the ASIC was to support the growing number of speech translations Google continually processes.

Currently, Google’s 3rd Generation TPUs are running in the market launched in 2018. TPU is built on a 28nm process, running at 700MHz and consumes 40W when running. TPU works by creating a grid of simplified ALUs. The data is sent via a PCIe bus to the grid. As the multiplication and addition of each vector are applied to each layer the result is passed to the next layer creating a pipelining effect throughout the 128x128 matrix of ALUs. Memory requirements are low as the output of one layer of ALUs is the input of the next layer. This also reduces power consumption as memory access is more power expensive than ALU computation.

GPUs from NVIDIA

Turing Architecture

NVIDIA has come up with GPUs based on Turing architecture to address fuelling growth in games and Deep Learning requirements. These GPUs are fine-tuned for high-performance computing in PC gaming, professional graphics, and Deep Learning inferencing.

Performance numbers for Quadro RTX6000 which is considered one of the superior GPUs is given below:

- 16.3 TFLOPS of peak single-precision (FP32) performance

- 32.6 TFLOPS of peak half-precision (FP16) performance
- 130.5 Tensor TFLOPS

Turing Tensor Cores along with continual improvements in TensorRT (Nvidia's Tensor Run Time Inferencing framework), CUDA, and CuDNN libraries, enable Turing GPUs to deliver outstanding performance for inferencing applications. They also add support for fast INT8 matrix operations to significantly accelerate inference throughput with minimal loss in accuracy.

Tensor Cores

New Tensor Cores are the key differentiating factors in Nvidia scaling up to the demands of Deep Learning. Tensor cores can deliver up to 125 Tensor TFLOPS for training and inference applications. These cores have been optimized and custom-built for Matrix-Matrix Multiplication operations which form the core of Neural Network training and inference. GV100 with CUDA delivers up to 1.8x higher performance than its predecessors.

Volta Tensor cores are also optimized to work with CUDA9 C++ APIs as they are accessible and exposed as Warp-level Matrix Operations. cuBLAS and cuDNN libraries have also been enhanced to provide new interfaces to utilize Tensor Cores for deep learning applications and networks.

Intel Hardware

Intel AI Research is pushing the limits of artificial intelligence and computing at every level, from atomic physics to datacentre orchestration. From hardware that excels at training massive, unstructured data sets, to extreme low-power silicon for on-device inference. Intel AI supports cloud service providers, enterprises, and research teams with a portfolio of multi-purpose, purpose-built, customizable, and application-specific hardware that turn model into reality.

Intel's ML Hardware Evolution

Intel Xeon Scalable processor

Intel Xeon processors provide an excellent platform for training deep learning models. The Intel Xeon Scalable processors can support up to 28 physical cores (56 threads) per socket (up to 8 sockets) at 2.50 GHz processor base frequency and 3.80 GHz max turbo frequency, and six memory channels with up to 1.5 TB of 2,666 MHz DDR4 memory.

Additional improvements include a 38.5 MB shared non-inclusive last-level cache (LLC or L3 cache), that is, memory reads fill directly to the L2 and not to both the L2 and L3, and 1MB of private L2 cache per core. The Intel Xeon Scalable processor core now includes the 512-bit wide Fused Multiply Add (FMA) instructions as part of the larger 512-bit wide vector engine with up to two 512-bit FMA units computing in parallel per core (previously introduced in the Intel Xeon Phi™ processor product line)1. This provides a significant performance boost over the previous 256-bit wide AVX2 instructions in the previous Intel Xeon processor v3 and v4 generations (formerly codenamed Haswell and Broadwell, respectively) for both training and inference workloads.

Intel Movidius

The Intel® Movidius™ Myriad™ X VPU is the third generation and most advanced VPU from Intel. It is the first of its class to feature the Neural Compute Engine — a dedicated hardware accelerator for deep neural network inferences. Interfacing directly with other key components via the intelligent memory fabric, the Neural Compute Engine is able to deliver industry-leading performance per watt without encountering common data flow bottlenecks encountered by other architectures. The Neural Compute Engine in conjunction with the 16 powerful SHAVE cores and an ultra-high throughput intelligent memory fabric makes Intel Movidius Myriad X the industry leader for on-device deep neural networks and computer vision applications. Intel's Myriad™ X VPU (Vision Processing Unit) has received additional upgrades to imaging and vision engines including additional programmable SHAVE cores, upgraded, and expanded vision accelerators.

Intel® Arria® 10 FPGAs

Intel FPGAs are blank, modifiable canvases. Their purpose and power can be easily adapted again and again for any number of workloads and a wide range of structured and unstructured data types. The Intel Vision Accelerator Design with Intel Arria 10 FPGA offers exceptional performance, flexibility, and scalability for deep-learning and computer-vision solutions from NVRs (Network Video Recorders) to edge deep-learning inference appliances to on-premises servers at a fraction of the cost and with significantly lower power requirements than most of the existing FPGA PCIe cards.

Intel® Nervana™ Neural Network Processors

Intel® Neural Network Processors (NNPs) was built from the ground up to breakthrough existing memory and data flow bottlenecks, enabling distributed learning algorithms and systems that will scale up deep learning reasoning, using more advanced forms of AI to go beyond the conversion of data into information-turning data into global knowledge.

Intel Nervana Neural Network Processor for Training

Intel Nervana NNP-T with inter-chip links (ICLs) is designed with a unique balance of computing, memory, and communications specifically to process DL workloads and move large amounts of data. NNP-T's high-speed ICL communications fabric enables customers to achieve near-linear scale by directly connecting NNP-T cards within servers, between servers, and inside and across racks, creating high-performance computing PODs. NNP-T maximizes processor and server POD utilization while reducing time to train, making it a highly energy-efficient alternative to general-purpose computing.

Intel Nervana Neural Network Processor for Inference

Intel® Nervana™ NNP-I was designed for intense, near-real-time, high-volume, low-latency compute. It can accommodate exponentially larger, more complex models and run dozens of models and networks in parallel. It was also designed for provide high inference throughput and power efficiency, plus programmable control for flexibility. Fully integrated voltage regulator (FIVR) technology optimizes SoC performance at different power envelopes for dynamic power management. On-die latest generation Intel® architecture (IA) cores that include Intel® Advanced Vector Extensions (Intel® AVX) and Vector Neural Network Instructions (VNNI) enable high levels of programmability so that AI practitioners can optimize for the next generation of models.

Software Requirement Specification

- CUDA Deep Neural Network (cuDNN) 7.5 library.
- NVIDIA CUDA 10.1.
- NVIDIA GPU driver 418.40.
- NVIDIA NCCL2 2.4.
- Anaconda 2018.12.

CHAPTER 6

DESIGN ANALYSIS

6. DESIGN & ANALYSIS

Mean-Face Analysis The results of the decoder are visualized in Figure 7. As one can see the key features of the original face are reconstructed. The wrinkles of old people are still present which indicates that the voice of older people is successfully detected as different and thus encoded as a feature. The same can be said for gender features as well as skin color. Non-important features for the voice like glasses are removed. In terms of age we start to see some limitations, because children start to look like young adults. We also notice some unexpected results. For some reason some people smile while others do not. We were not able to link these smiles to any voice feature like a certain pitch and further testing would be required. Even more surprising is how beards are treated. One would assume that beards do not affect a persons voice and yet beards remain clearly visible. They often get reduced but never completely vanish. Related to this is the last image in Figure 7 where the brown skin color was lost and replaced with white skin while the beard remained very visibly. Again we were not able to link a beard to a voice feature.

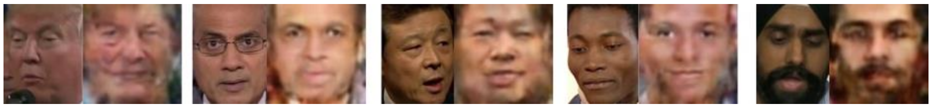


Fig. 7. Left is the ground truth and right is the reconstructed face.

Lip2Speech

We use LRW dataset for evaluating our model. We sample 153 videos from the whole dataset with people of different age, gender, ethnicity and accent. In fig 8, sample faces and the corresponding generated melspectograms are shown. We show Short-time Objective Intelligibility (STOI), extended STOI (ESTOI), Per- ceptual Evaluation of Speech Quality (PESQ), and Word Error Rate (WER) metrics results for the generated speech in table 2. STOI, ESTOI, and PESQ metrics measure the speech intelligibility which is the the degree to which speech sounds (whether conversational or communication-system output) can be cor- rectly identified and understood by listeners.

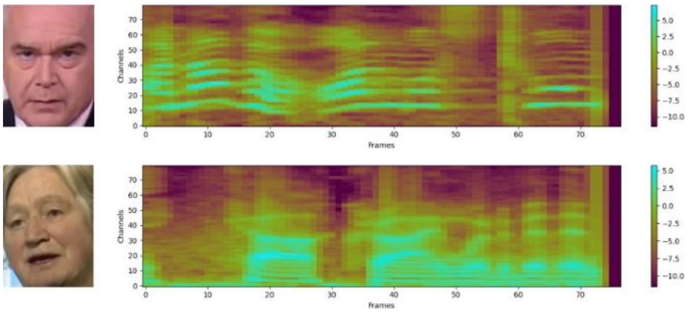


Fig. 8. Faces and the melspectograms generated with our Lip2Speech network.

CHAPTER 7

IMPLEMENTATION

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 it constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods a 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.

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 theobjectives have been met. Errors are noted down and corrected immediately.

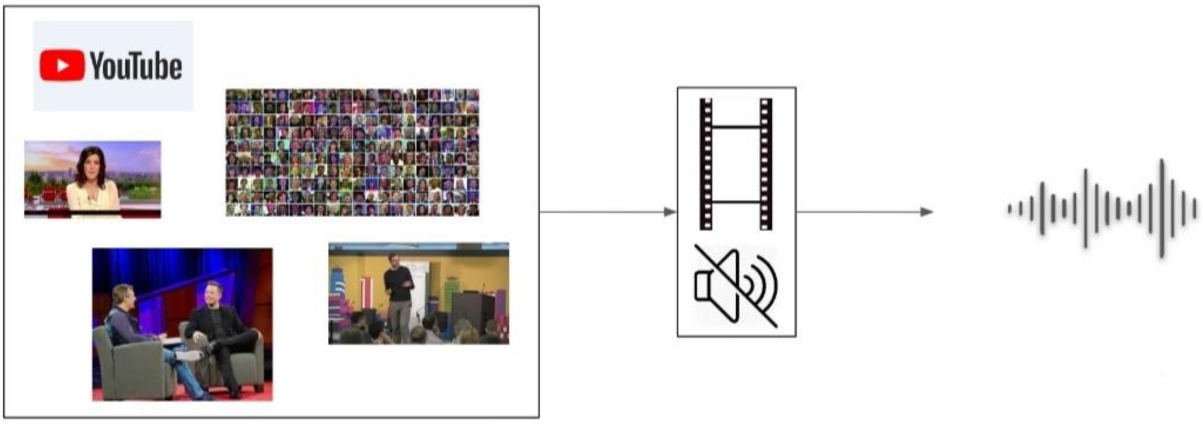
2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.

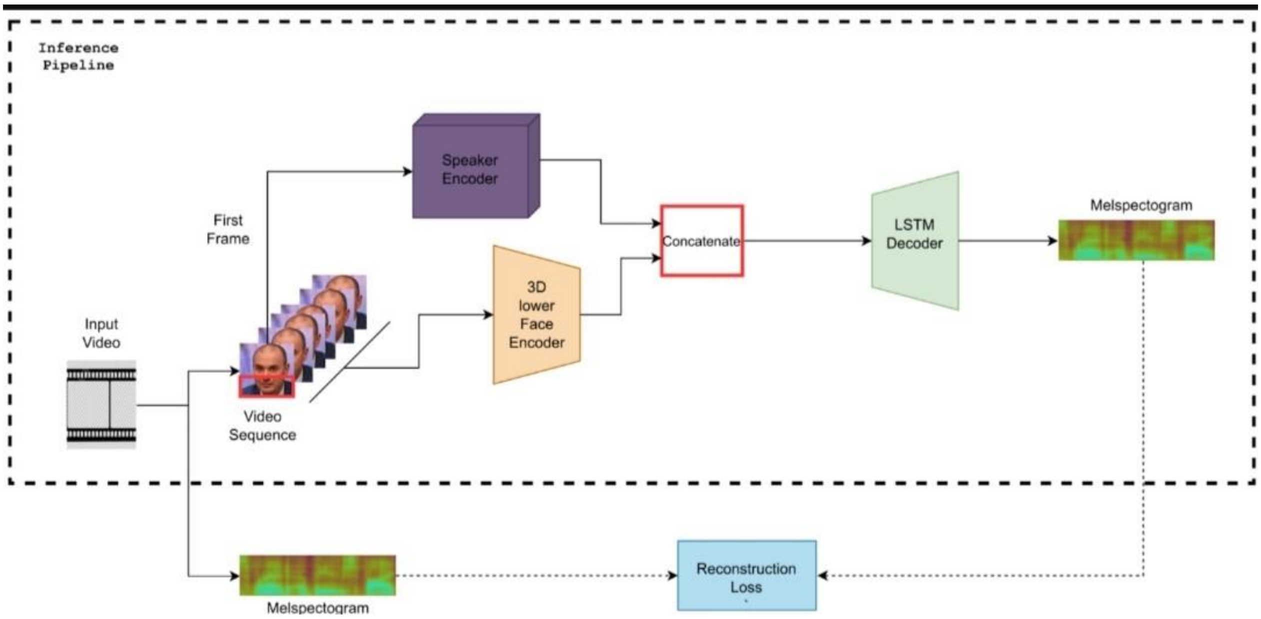
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

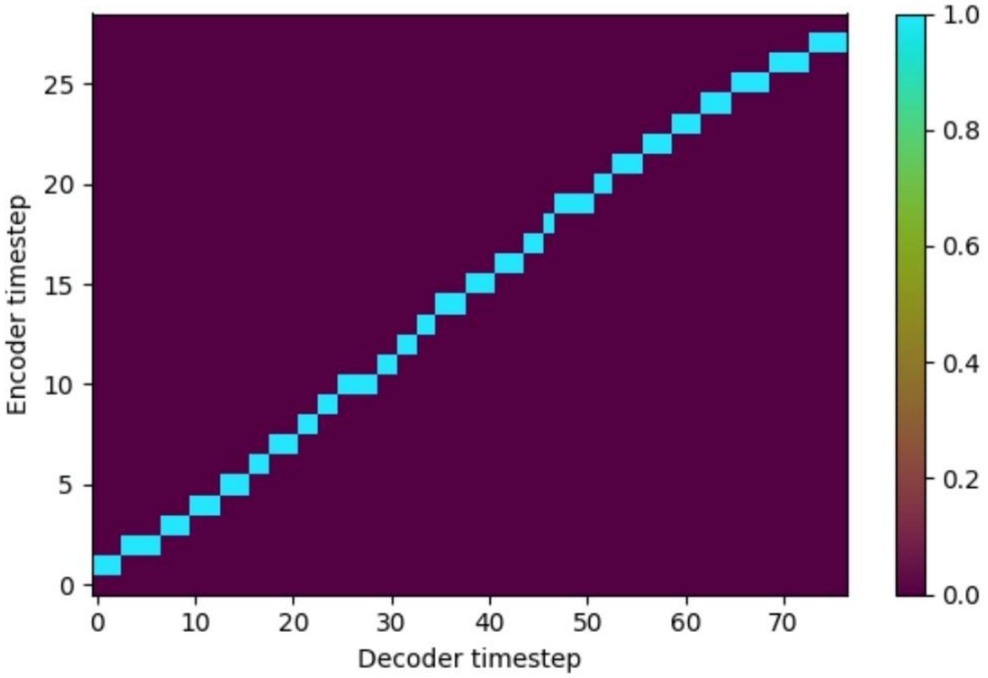
CHAPTER 8

SNAPSHOTS

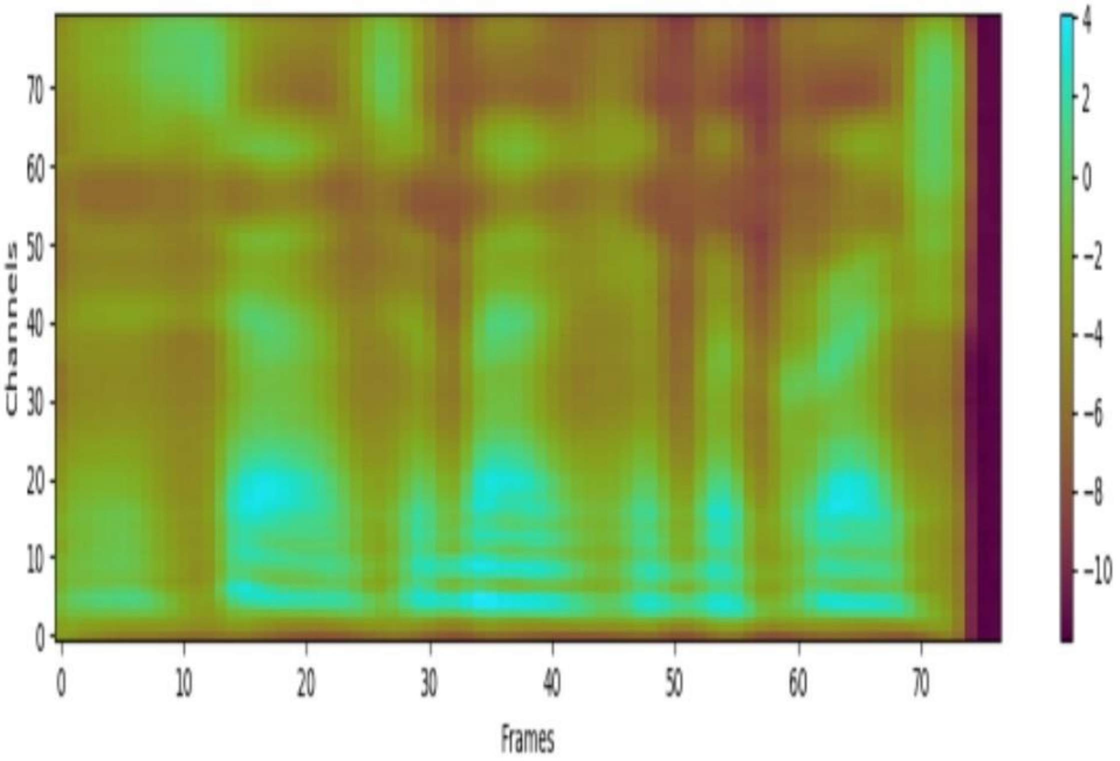
8. SNAPSHOTS


YouTube


Inference
Pipeline
Speaker
Encoder
First
Frame
Concatenate
LSTM
Decoder
Melspectogram
3D
lower
Face
Encoder
Input
Video
Video
Sequence
Melspectogram
Reconstruction
Loss



Encoder timestep
Decoder timestep
0
5
10
15
20
25
0
10
20
30
40
50
60
70
0.0
0.2
0.4
0.6
0.8
1.0



Channels
Frames
0
10
20
30
40
50
60
70
0
10
20
30
40
50
60
70
4
2
0
−2
−4
−6
−8
−10

CHAPTER 9

CONCLUTION

9. CONCLUTION

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project:

- Automation of the entire system improves the efficiency
- It provides a friendly graphical user interface which proves to be better when compared to the existing system.
- It gives appropriate access to the authorized users depending on their permissions.
- It effectively overcomes the delay in communications.
- Updating of information becomes so easier
- System security, data security and reliability are the striking features.
- The System has adequate scope for modification in future if it is necessary.

10. REFERENCE

- https://doi.org/10.1121/1.5042758
- http://dx.doi.org/10.1109/CVPR.2017.367
- http://dx.doi.org/ 10.1145/3197517.3201357
- https://www.sciencedirect.com/science/article/pii/S0960982203006638
- http://papers.neurips.cc/paper/ 9015-pytorch-an-imperative-style-high-performance-deep-learning-library.