Sensor-based Bio-signal Processing and Modeling for Human Behavior and Expression Recognition

UNDERGRADUATE THESIS

Submitted in partial fulfillment of the requirements of
BITS F421T Thesis

By

V.A.S ABHINAV
ID No. 2018B1A70979G

Under the supervision of:

Dr. Prof. Tanja SCHULTZ
&
Dr. Prof. Sougata SEN

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, GOA CAMPUS
May 2023

Declaration of Authorship

I, V.A.S ABHINAV, declare that this Undergraduate Thesis titled, 'Sensor-based Bio-signal Processing and Modeling for Human Behavior and Expression Recognition' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Abhinav

Date: 12.05.23

Certificate

This is to certify that the thesis entitled, "*Sensor-based Bio-signal Processing and Modeling for Human Behavior and Expression Recognition*" and submitted by V.A.S ABHINAV ID No. 2018B1A70979G in partial fulfillment of the requirements of BITS F421T Thesis embodies the work done by him under my supervision.

Supervisor	*Co-Supervisor*
Dr. Prof. Tanja Schultz	Dr. Prof. Sougata Sen
Principal Investigator,CSL,	Asst. Professor,
University of Bremen	BITS-Pilani Goa Campus
Date: 12.05.23	Date:

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, GOA CAMPUS

Abstract

Bachelor of Engineering (Hons.)

Sensor-based Bio-signal Processing and Modeling for Human Behavior and Expression Recognition

by V.A.S ABHINAV

Human facial expressions play a pivotal role in conveying meaningful thoughts during a conversation. FACS(Facial Action Coding System) provides a framework to quantitatively and qualitatively describe facial expressions. Action units form the basis of this framework. In this project, several Random Forest Models are trained which leverage fEMG data to predict action units. An accuracy of about 80% is obtained with the best-performing model. An effort was made to analyze features which played a prominent role in differentiating between the samples. A script was written to get real-time output of action units using the models trained which was streamed into blender visualization software. It was observed that the model was good at picking up action units nearer to the lips compared to the others. This provides empirical evidence that fEMG can be leveraged as a suitable modality in action unit recognition studies. Future improvements are planned to improve upon the performance of models using fEMG-based action unit recognition.

Acknowledgements

I would like to express my sincere gratitude to Dr. Dennis Küster, who provided invaluable guidance and support throughout my research. His expertise, insights, and constructive feedback have been instrumental in shaping this thesis and improving its quality.

I am also grateful to the members of my thesis committee, Dr. Prof. Tanja Schultz and Dr. Felix Putze, for their valuable comments, feedback, and suggestions. Their critical evaluation and constructive criticism have greatly helped me to improve my work.

I would like to acknowledge the support and assistance of Cognitive Systems Lab at Uni Bremen, who provided financial support for my research. This funding has enabled me to carry out my research and complete this thesis.

I would also like to extend my appreciation to Dr Sougata Sen, Dr Rainer Koschke and Marcel Steinbeck, who provided valuable feedback, suggestions, and encouragement throughout my research. Their constructive criticism and support have been a source of motivation and inspiration for me.

Lastly ,I would like to thank my family and friends for their unwavering support, encouragement, and patience throughout my academic journey. Their love and support have been a constant source of motivation for me.

Thank you all for your invaluable support and encouragement.

Contents

List of Figures

List of Tables

Abbreviations

fEMG	**f**acial **E**lectro**M**yo**G**raphy
VR	**V**irtual **R**eality
AR	**A**ugmented **R**eality
FER	**F**acial **E**xpression **R**ecognition
AUR	**A**ction **U**nit **R**ecognition
FACS	**F**acial **A**ction **C**oding **S**ystem
ADFES	**A**msterdam **D**ynamic**F**acial **E**xpression **S**et
LSL	**L**ab **S**treaming **Layer**
AU	**A**ction **U**nit
CERT	**C**omputer **E**expression **R**ecognition **T**oolbox
SEE	**S**oftware **E**ngineering **E**xperience
HMD	**H**ead **M**ounted **D**isplay

Chapter 1

Introduction

1.1 Background

User presence in collaborative settings in virtual worlds is generally achieved by avatars [23] [14]. In these settings, people are not in the same room and, thus, cannot see each other or interact with one another to discuss, solve problems, or simply to meet and greet. A large part of human communication is non-verbal and visual. However, this information is lost when only the voice is transmitted as is the case in most VR settings.

The human face in particular plays a major part in expressing emotions that conveys important information. Hence it becomes imperative that a system be designed which can capture this crucial information from the human face. One such VR platform which is trying to solve this problem is SEE [1] (Software Engineering Experience). SEE 1.1 is a platform developed by the research group led by Rainer Koschke offering virtual rooms that can be entered by participants by way of various devices, such as regular desktop computers and VR/AR hardware. The participants can see not only shared content but also other team members, too, and talk to each other through a voice channel. It thus creates a more natural user experience. Current implementations of SEE use HMD(Head Mounted Displays) and *HTC Vive Facial Tracker* to capture non verbal information. However, if the participants enter the virtual world by wearing head-mounted displays (HMD), their face is mostly covered and cannot be captured by cameras. Trackers such as the HTC Vive Facial Tracker can observe only the lower region of the face, not covered by an HMD. Moreover, in many workplaces and everyday situations, cameras cannot or must not be used. In such situations, other indicators of emotions must be utilized as a proxy,

[1] https://see.uni-bremen.de/

for instance, the sensors attached to the face.

Hence this project will research, develop, and evaluate an approach that captures emotions from affective prosody i.e. facial electromyography (fEMG) of a person to create facial expressions for their avatar showing the captured emotions authentically.This approach enables humans participating in virtual worlds,by way of avatars, to leverage a richer set of means of non-verbal communication.Through the findings and knowledge gained through this thesis project, the ultimate aim is to capture emotional prosody in VR and desktop environments and visualize the identified emotions via facial expressions of the avatars in real-time in SEE(Software Engineering Experience).

The potential applications are manifold.For instance,elderly people can meet their family and friends living remotely or interact with others taking care of their health. Engineers in distributed organizations can meet in virtual workplaces interacting with both their virtualized engineering products and their collaborators in ways that are much more effective than today's video-conference and screen-sharing tools allow.Teachers can meet their students in a more engaging virtual class room

Figure 1.1: SEE environment

1.2 Objectives

This thesis project strives to utilize the fEMG modality to create a strong foundation for detecting facial expressions. To sum up, the main goal and questions addressed in this thesis are

- **RQ1:**How can we accurately map the facial expressions detected and display them in real-time if possible, on a VR avatar's face?
  - This comprises identifying the correct framework to detect facial expressions.
  - Constructing a relevant ML model which can accurately detect the facial expressions with the framework chosen in real time.

- **RQ2:**Identify the necessary components of the pipeline for translation of the facial expression data to the visualization software.
  - This comprises identifying the necessary steps for animating the predictions of facial expression in a visualization software.

1.3 Structure of report

Since I aim to construct an ML model which can predict facial expressions leveraging fEMG data. I would first need to identify of how I am going to predict the facial expressions. This means that I need to have a suitable encoding system in place for facial expressions. Once I have such a system in place, I can construct a dataset from which my ML model can learn. To construct such a dataset I would need a coherent plan to record fEMG data with correct labels which can be used for model training. The trained model can be then used to evaluate the effectiveness of fEMG as a modality for facial expression prediction. Hence the thesis report is divided into 3 sections. Section 1 i.e Chapter 2 goes into the existing literature in the facial expression recognition domain. It also gives a brief overview of the current facial expression animation tools that presently being used by researchers and finally ends with the current existing gaps within the research which are explored in this thesis. Section 2 i.e Chapter 3 goes in depth about the adopted methodology for the proposed objectives. This consists of explanation about the experimental setup,synchronization,data preparation methods and finally the model selection process. Experimental setup describes the data collection process that was undertaken. Synchronization section explains how the fEMG data was matched with the right labels. Data preparation delves into the adopted method for constructing the datasets for model training.

Chapter 4 presents results relating to the model and the outputs obtained in real time using the models trained. Particular focus is paid to the variety of information that can be potentially captured using fEMG as a modality.

Chapter 2

Literature Review

Emotions are an integral part of everyday human experience. In order to fully understand each other, we need to consider the emotional state of others. Human facial expressions are one essential aspect which can convey the emotional state of a person. Facial Emotion Recognition (FER) is the technology that analyses facial expressions from both static images and videos in order to reveal information on one's emotional state. Action unit recognition (AUR) is an important research direction within FER used for predicting facial expressions. AUR aims to automatically identify movements of facial muscles 2.1 that correspond to specific emotions, expressions, and actions. This approach analyzes the dynamics of subtle changes in the face, such as wrinkling of the nose, raising of the eyebrows or lip corner pulling, to determine which emotions are being displayed. The major challenges of AUR include dealing with variability the in facial expression [4], individual differences, head motion, and occlusions [12]. Therefore, designing robust and accurate algorithms for AUR remains an active area of research[2]. Traditional approaches to action unit recognition were based on manual recognition by people who had undergone laborious training in the FACS (Facial Action Coding System) developed by Ekman et al.(2002) [8]. This process was very time consuming much to the detriment of experimental studies. Research soon began upon trying to replace human input with automated computational approaches for action unit recognition. [15] CERT (Computer Expression Recognition Toolbox) was one of the first widely developed tools which was used for action unit classification. This tool was is now offered as a state of the art classifier as part of the paid iMotions[1] software.

[1] https://imotions.com/

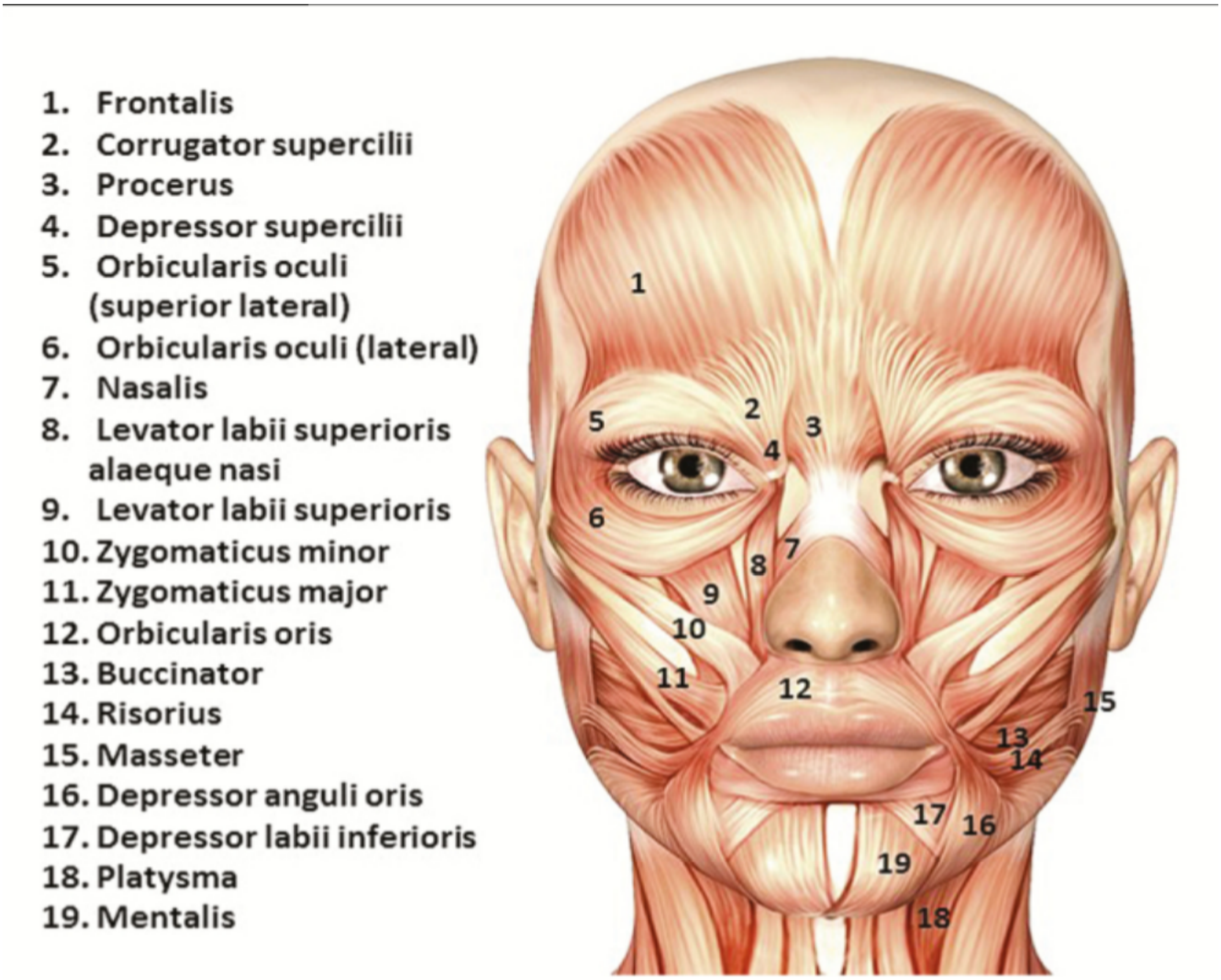

FIGURE 2.1: All facial muscles [17]

Another paid software existing in the market is the FaceReader [2] package offered by Noldus. None of these software packages are accessible to the open source community which makes it difficult for research studies to be conducted using action unit classifiers. This gap was filled by OpenFace[2], which offered a lot of features such as face detection,dace landmark detection etc. including action unit classification as a part of its software.

Up until now we have only focused on how state of the art action unit recognition is done. But how do we convert these action units detected into meaningful visualizations on avatars?. Facial expression visualization software can help us in gaining deeper insights into human emotions and behaviors. One such tool is FACSHuman [10]. The application is a resourceful instrument that

[2] https://www.noldus.com/facereader

generates photos and videos of facial expressions by employing the FACS [8]. It is based on the Makehuman[6] software program and consists of three additional plugins for Makehuman. The first FACSHuman plugin developed, which is presented here, allows complex facial expressions to be created, and the elements of muscular movements of the face, skin, and eyes to be defined as well as those of the jaw and the head. The second plugin FACSAnimation (FANT) is a plugin for creating animations of facial expressions. The user can compose, create, and record animations by direct creation or by mixing different expressions created in the facial expression creation tool. The third and last plugin FACSSceneEditor (FSCE) defines the lighting of a scene. Another tool for visualizing animations is FACSvatar. FACSvatar [22] is a state-of-the-art software program that uses the FACS to generate realistic computer avatars that can accurately mimic and respond to human facial expressions. The program can analyze and recreate facial expressions with remarkable accuracy and precision, making it a valuable tool for studying human behavior and emotion. By creating avatars that can accurately portray and respond to human emotions, FACSvatar has the potential to enhance social and emotional connections in virtual environments and contribute to a better understanding of human behavior and emotion in various fields of study. The technology behind FACSvatar represents a significant step forward in the development of more realistic and responsive virtual environments, and its potential applications are broad and far-reaching especially in the context of collaborative environments which involve communication as in the aforementioned SEE[3] code city environment developed by Rainer Koschke's group.

Considerable studies [18],[19] have already been conducted for validating the authenticity of action units produced by the software primarily focusing on the video input to classifiers. To the best of my knowledge, there is yet a gap in research which concerns action unit output based on other modalities like fEMG data, which will be addressed in the thesis.

[3]https://see.uni-bremen.de/

Chapter 3

Methodology

3.1 Proposed Framework

The proposed framework comprises of creating a dataset which contains synchronized video modality data with fEMG recordings and the output labels corresponding to suitable action units. An ML model will be trained using this data, and then the predictor will be constructed in such a way that action units are output only based on the incoming EMG data. The framework can be visualized with the help of the flowchart 3.1 shown below. The flowchart can be broken down into the following list of tasks. Following sections will go into more details regarding the tasks.

- Task 1: Designing our Experimental Setup (Experimental Phase)
- Task 2: Designing and Training our ML model (Training Phase)
- Task 3: Evaluating our model (Execution phase)

3.2 Task 1: Experimental Setup

To train an ML model which is good at recognising action units from the fEMG data, it is necessary to get the right output labels for some corresponding action whilst also keeping in mind the that ground truth label is correct and based on widely accepted standards i.e based on the FACS system. One of the ways would be to ask the subject to perform some predesignated

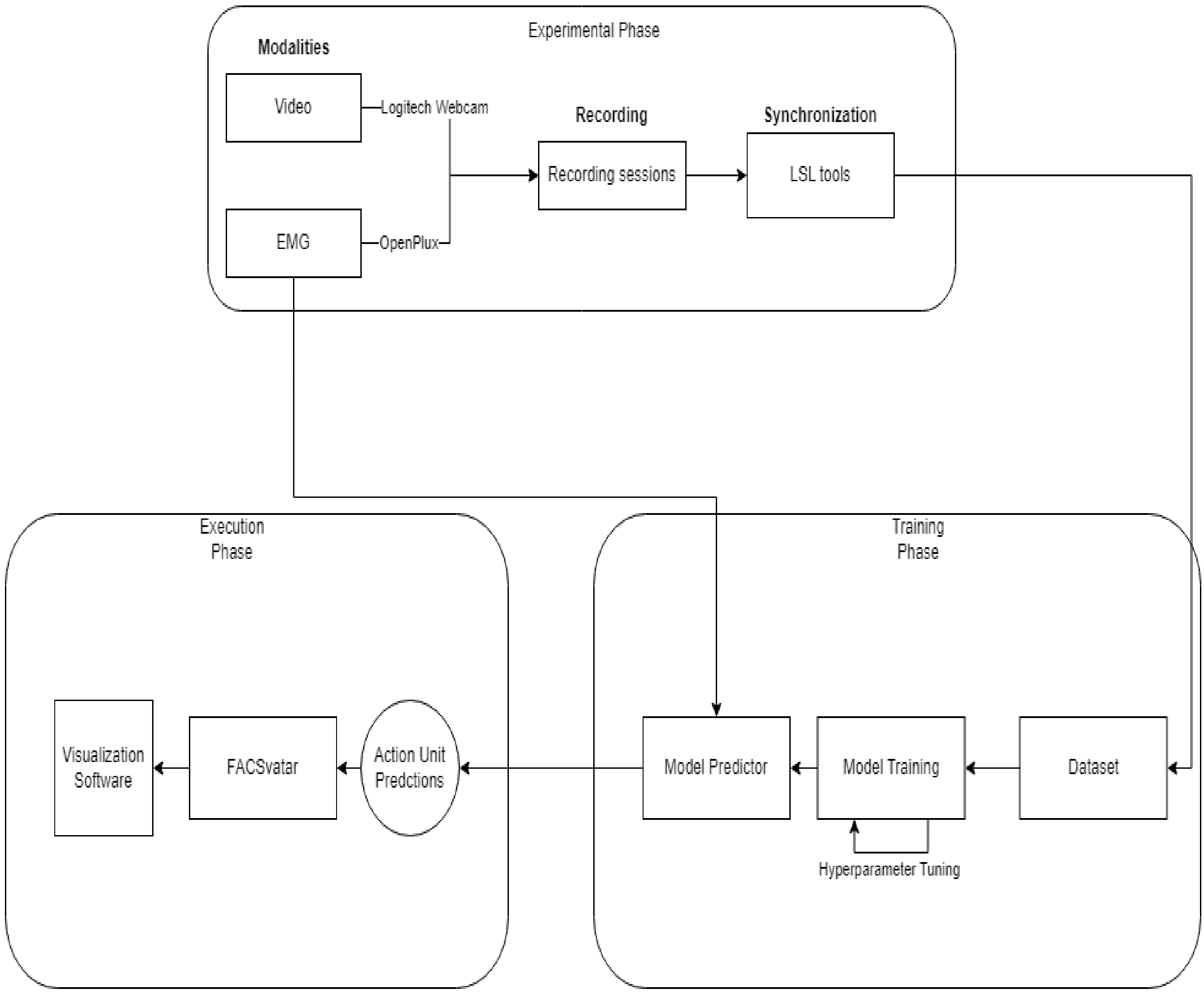

Figure 3.1: Flowchart for displaying framework of thesis work

action and manually get the labels from a human FACS coder on a frame by frame basis , but this would be a rather time-consuming process. Instead, I decided to manually tag the labels for each action with the correct ground truth label by asking the subject to perform an action that corresponds to the correct ground truth label. But this would place a stress on how good or accurately the action is performed during the data acquisition process. One way to circumvent this issue would be to leverage the use of pre-existing software which gives the action unit labels in real time which can be used to give feedback to the subjects to correct their action. This should help them in performing an action which correctly maps to the intended output. This puts forward another issue of deciding which action unit classifier would be the best one to use. Three action unit recognition software were considered as possible options.

1. FACET by Imotions
2. FaceReader by Noldus
3. OpenFace (Opensource)

3.2.1 Choosing the best Action Unit Recognition software

Choosing the best action unit recognition software from the aforementioned three needs a proper framework to be constructed for comparison of these three softwares.FACET,FaceReader and OpenFace all are action unit recognition softwares which are based on Video Modality. Each has an associated predictor which gives intensity predictions for some action units present on the face. FACSvatar developed by Struijk et al provides a modular framework for converting facial expression to procedural avatar animation based on FACS system. Figure 3.2 describes the framework of the FACSvatar. FACSvatar was developed by the authors to make use of the opensource library OpenFace, however, within this project, I was able to do minor tweaks to the FACSvatar software to make it compatible to be used with Imotions and FaceReader.

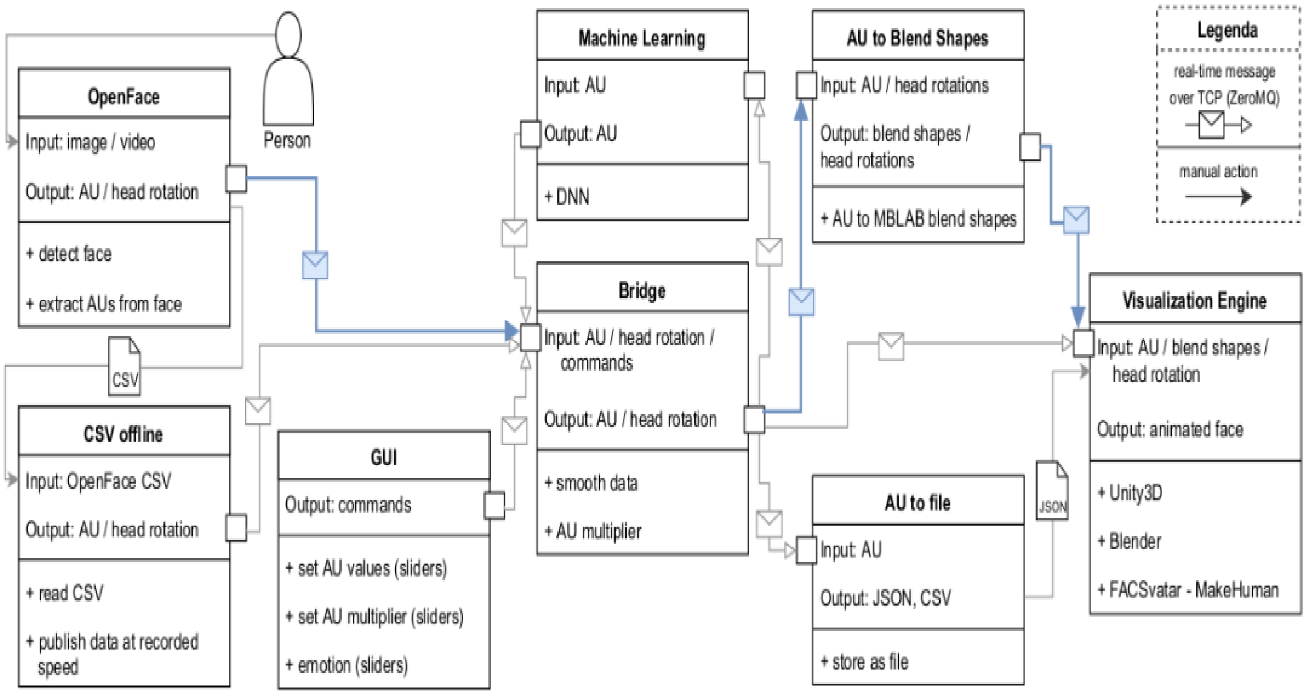

FIGURE 3.2: FACSvatar framework [22]

Sr no	Video Title
1.	medit F07-Anger-Face Forward
2.	medit F07-Contempt-Face Forward
3.	medit F07-Disgust-Face Forward
4.	medit F07-Embarrass-Face Forward
5.	medit F07-Joy-Face Forward
6.	medit F07-Neutral-Face-Forward
7.	medit F07-Sadness-Face-Forward

TABLE 3.1: Videos Chosen for comparison

All three software were run on the videos from the ADFES(Amsterdam Dynamic Facial Expression Set data set. The chosen videos are displayed in the table 3.1. The softwares provided predictions of action units on a frame by frame basis. A script was written to visualize the action units in real-time while the ADFES videos were run on loop. On close visual inspection it was seen that FACET by Imotions was the best software. Figure 3.3 provides an example of

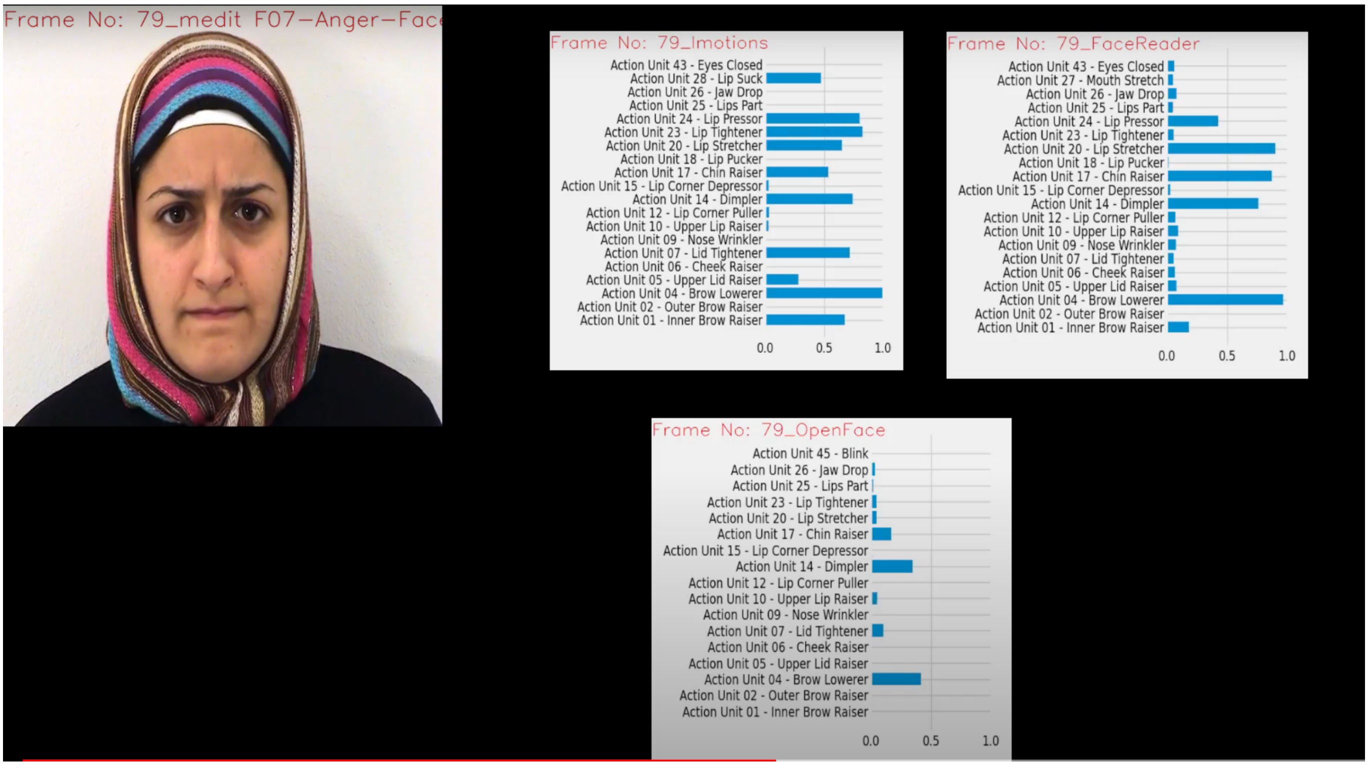

FIGURE 3.3: Visual Inspection of Various AU recognition software

how the visual inspection was carried out. Hence Imotions was chosen as the classifier of choice. The comparison videos are attached in Appendix A. Although these models by themselves are not foolproof and tend to be erroneous but they were deemed good enough to give a feasible output for the intended purposes.

Apart from this I was also able to visualize the action units output by different software in a blender visualization system using the MB-Lab plugin for blender. Blender was used as the visualization system of choice for its easy functionality of rendering animations which can be stored as well as FACSvatar providing support for streaming into blender. This also consolidated our belief that Imotions provides the best output out of all the three. Figure 3.4 provides one of the visualization examples. The comparison videos with blender are attached in the Appendix A.

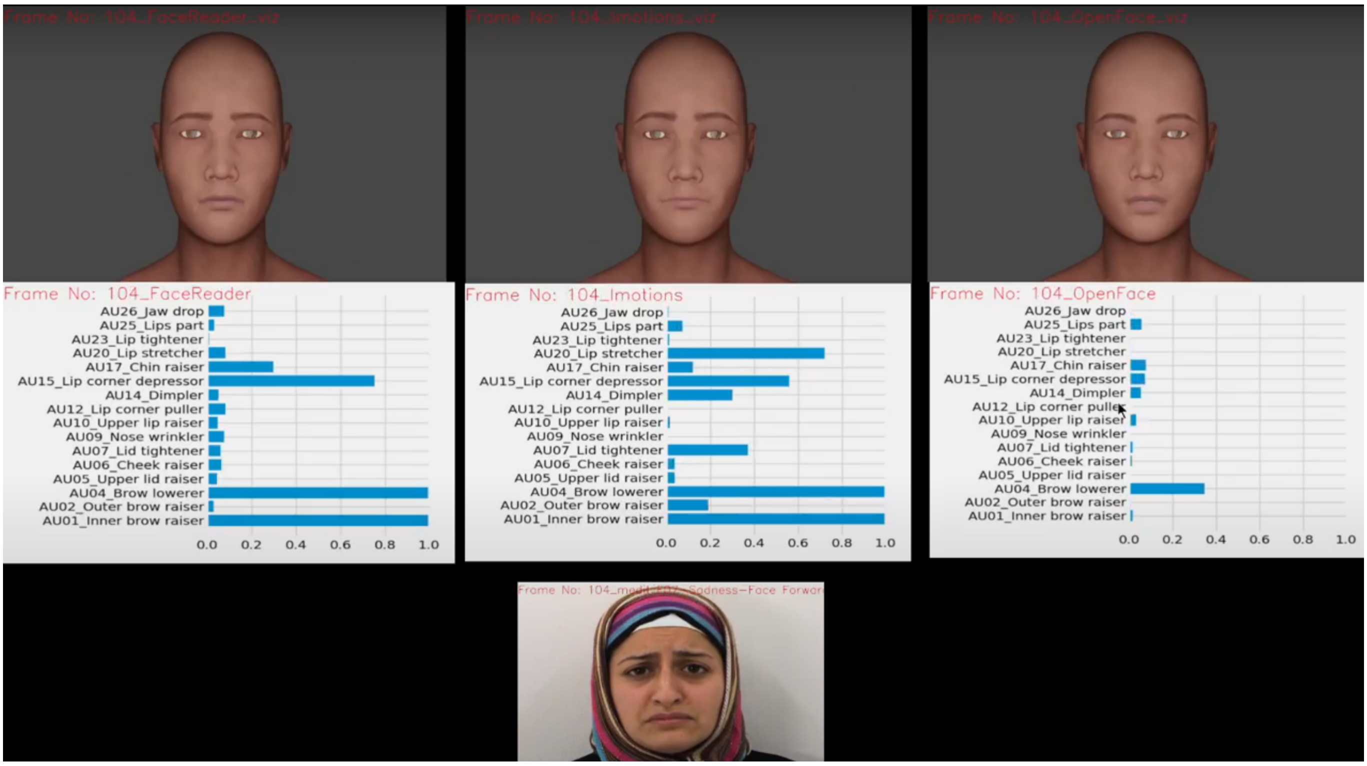

FIGURE 3.4: Visual Inspection of Various AU recognition software in Blender

3.2.2 Recording Setup

The recording setup consists of one host computer, connected to one USB camera. The subject will be seated in front of the stimuli display screen approximately 70cm away. 4 facial electrodes positions were chosen for recording. The fEMG setup is a bipolar recording setup consisting of 4 channels listed in Table 3.2 and showcased in Figure 3.6. Each channel is placed in a specific position on the face. These positions are defined by the internationally accepted guidelines in Fridlund and Cacioppo 1986 [9].Figure 3.7 gives the visual representation of all positions specified in the guidelines. It is known that the fEMG signals occur in the range of 15-500Hz [5], so the sampling frequency was chosen to be 2000 Hz as required by the limitations imposed by the Nquist Theorem which states that a periodic signal must be sampled at more than twice the highest frequency component of the signal.

The recording protocol can be divided into two phases.

First Phase - Training

Before the fEMG electrodes are attached to the participant's face. The participant can train himself to mimic the expressions corresponding to target action unit labels with the training script provided. Whenever the training script is initiated. Training instructions 3.8 are provided at the start. Once the participant clicks on "OK" a new window is displayed. This window 3.9 consists of three parts. The first part displays the stimulus video of a particular action unit which the participant is expected to imitate. Second part displays the webcam feed of the participant itself. The third part shows a trackbar graph in the range of 0-1 which acts as a feedback loop for the participant. The participant is asked to imitate the expression shown in the stimulus window. If the imitated expression is good enough, the track bar should display a high rating. In this way the participant can practice and improve their imitations. Once they feel confident in their imitation. They can move on the second phase.

Second Phase-Recording Now that the participant is familiar with the imitations,we need to record our participant's data to construct our datasets. Hence the second phase corresponds to recording trials. Each recording trial consists of the subject being shown stimuli pertaining to a particular action unit in sequence. The stimulus videos chosen are from MPI Video Database[13]. The stimulus videos consist of a person trained in FACS enacting specific action units as showcased in 3.5. These specific action units will become our ground truth for fEMG frames. The trial starts with the display of a window 3.11 which provides instructions to the participant regarding the session. Once the participant clicks on "OK". A white screen is displayed for about 5 seconds. During this period, they are expected to keep a neutral expression. In experimental terms, this corresponds to baseline recording. Once the baseline period is over, a new screen is displayed 3.12. This screen consists of 2 sub windows in a grid. The first window shows the stimuli window. The second window shows the webcam feed of the subject itself , where they can verify their imitation again. Once they are confident, they should click on the record button. The participants are now asked to do the expression at peak intensity and click on the

Sr no	Electrode Position
1.	Corrugator Supercilii
2.	Orbicularis Oculi
3.	Zygomaticus Major
4.	Levator Labii

TABLE 3.2: Electrode Positions

middle mouse button as well as hold down the push button provided to them. Once this is done the recording starts, where the participants hold the expression for about 5 seconds. Similar procedure is repeated for the rest of the action units.50 recording trials were performed in total. During this entire procedure the fEMG signals are being recorded using EMG sensors. The Stimulus Videos are linked to in Appendix A.

FIGURE 3.5: Snapshot of stimuli that will shown to the participant [9]

3.2.3 Synchronization

The synchronization of the data recorded is handled using the Lab Streaming Layer (LSL) tools. During the experimental sessions, the recording program sends the frame counts as well as the corresponding action unit being recorded to the LSL client which is set up. The same LSL client

FIGURE 3.6: Subject Electrode Placement

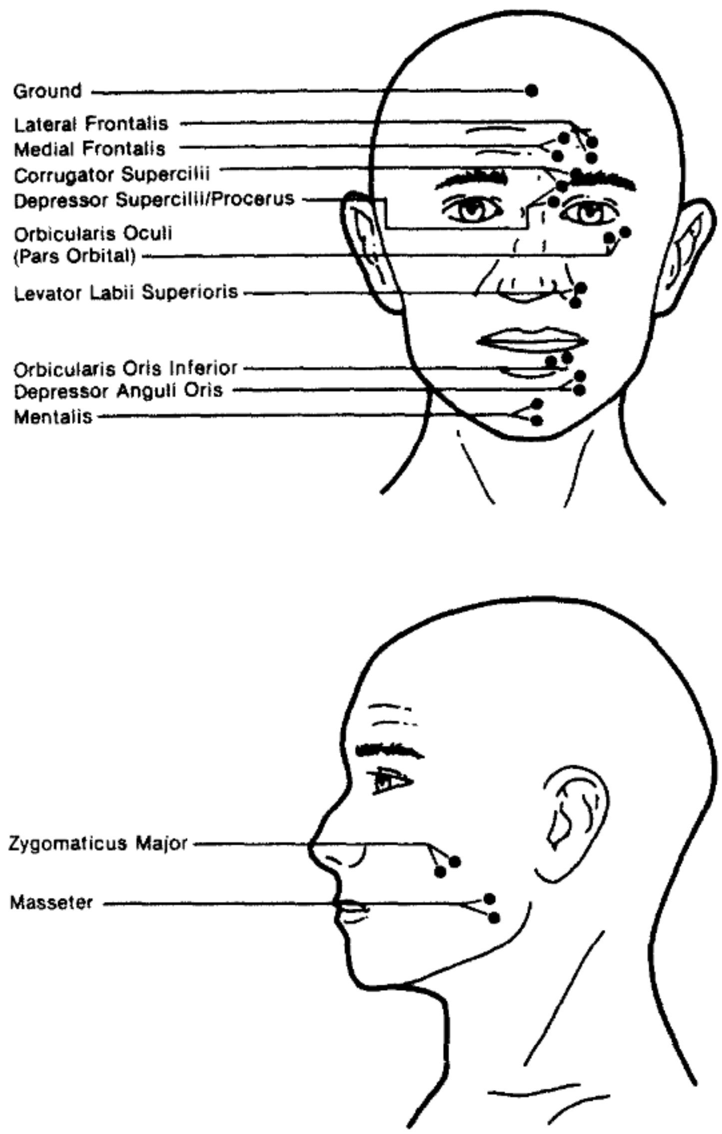

FIGURE 3.7: fEMG position atlas [9]

also listens to the EMG data stream. Once the recording is done. The recorded data is stored as .xdf file per trial. During the post-recording analysis of the data using xdf file readers which corrects the time offsets for each data stream using in-built algorithms i.e putting each data stream on the same clock.

The timestamps corresponding to frame counts and action units are used to extract the EMG data. Consider for example you want to extract the femg data between frame count **i** and frame count **i+1** for some particular action unit. Let the timestamps corresponding to both the frames be **t1** and **t2**. All fEMG data samples whose timestamps fall between **t1 & t2** are mapped to frame i. The following technique is applied for all frame counts, and an EMG mapping is produced. To ensure that this mapping is correct , the push button values recorded during the

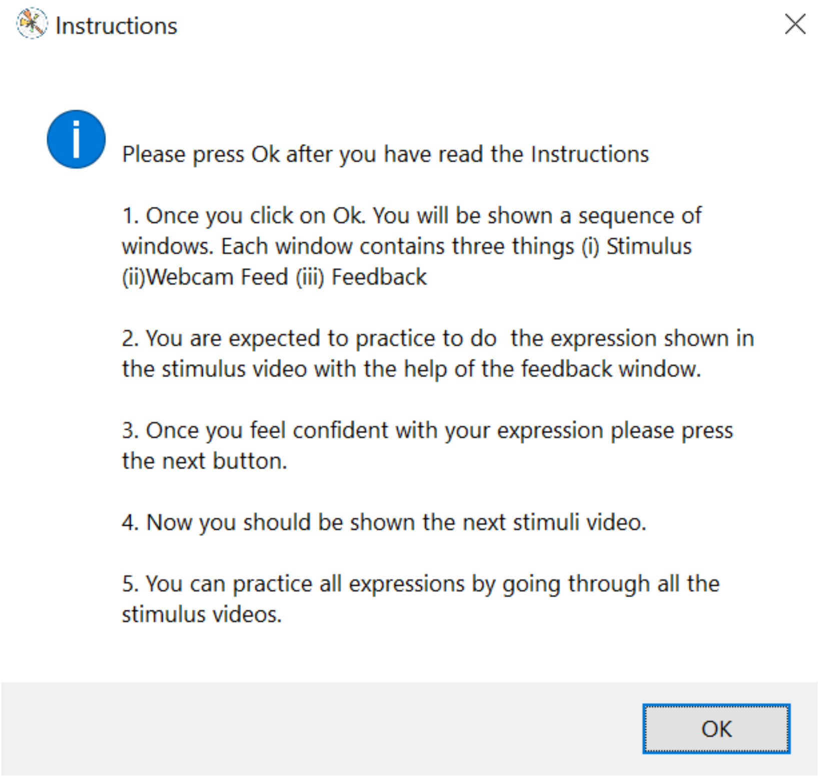

Figure 3.8: Training Instructions

trials are also used as a filter to produce the correct mapping. This mapping is then used to construct the EMG dataset and is stored in a csv format. Figure 3.10 showcases the idea.

3.2.4 Pre-Processing Pipeline

Raw fEMG data itself consists of a lot of artifacts which need to be removed. Some researchers prefer to work with the raw data itself. and run their feature extractors directly on raw fEMG. However, it is still advisable to attenuate the noise components.
Any fEMG signal being recorded is generally contaminated by following sources of noise [20].

1. **Inherent Noise in electrical equipment**: All electronics equipment generate noise. This cannot be eliminated but only reduced using high grade equipment.
2. **Ambient Noise**: Electromagnetic radiation is the source of this noise. PLI(Power Line Interference) falls in this category. It is generally of the frequency 50Hz
3. **Motion Artifact**: This can occur due to the movement between the electrode interface and movements of the cable. Motion artifacts fall in the range of 0-15Hz

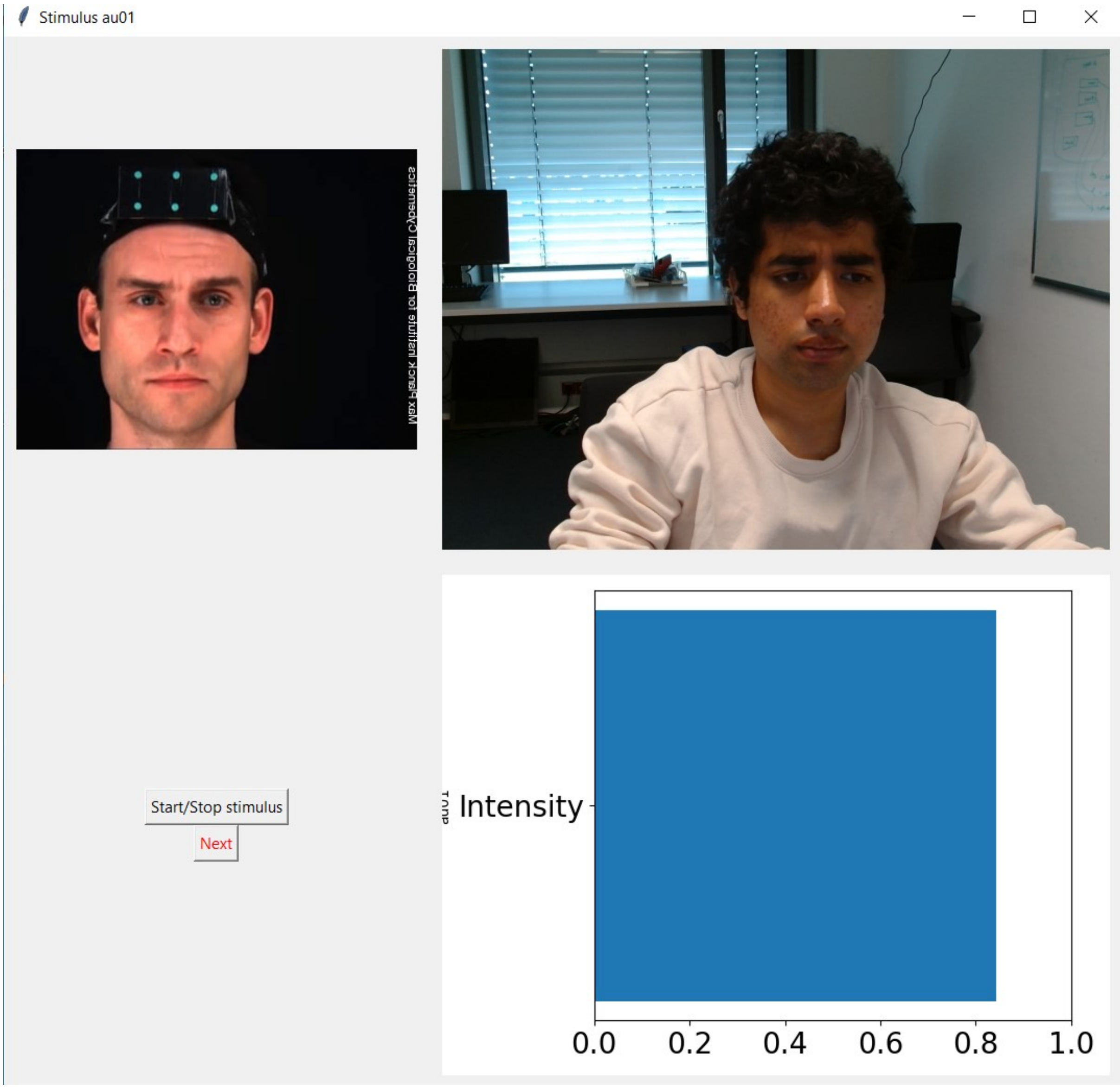

FIGURE 3.9: Trainer provided to participants to practice their imitations

Within this project an effort is made to study the effects on the model performance based on the data being pre-proceessed or otherwise. Following steps were taken in the pre-processing of data.

1. **Segmentation**

   The EMG signal is broken down into disjoint segments of specified length in milliseconds.

2. **Notch Filter**

   A notch filter was applied to remove PLI(Power Line interference) of 50Hz.

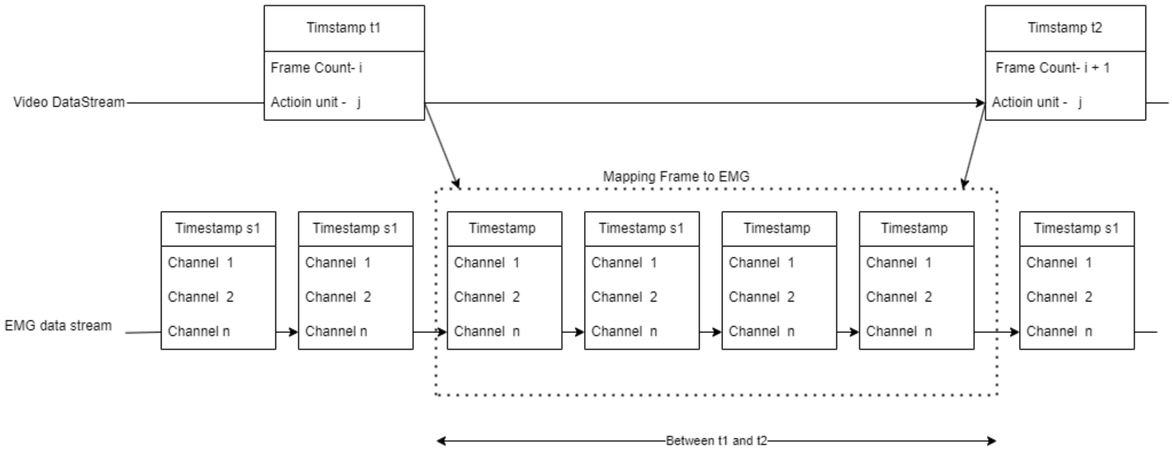

Figure 3.10: Mapping frame to EMG samples

3. **Band Pass Filter**

   A Butter-worth 4th order filter was applied to to filter out frequencies in the range of 15Hz to 500Hz.

4. **Full Wave Rectifier**

   Converts channels values to absolute values.

3.3 Task 2: Training an ML Model

Random forest model will be used as the initial model for prediction.The basic idea behind a random forest is to create a large number of decision trees, each of which is trained on a random subset of the data and a random subset of the features. This randomness helps to prevent over fitting, which can occur when a model becomes too complex and learns to fit the noise in the data rather than the underlying patterns.One of the advantages of a random forest is that it can handle high-dimensional data, noisy data, and data with missing values. It is also relatively easy to interpret, since it can provide information about which features are most important for making predictions.

3.3.1 Features

Training an ML model requires the right set of features to be fed to the Model for optimal performance. Considering we want our model to work in real time in the future. It is important

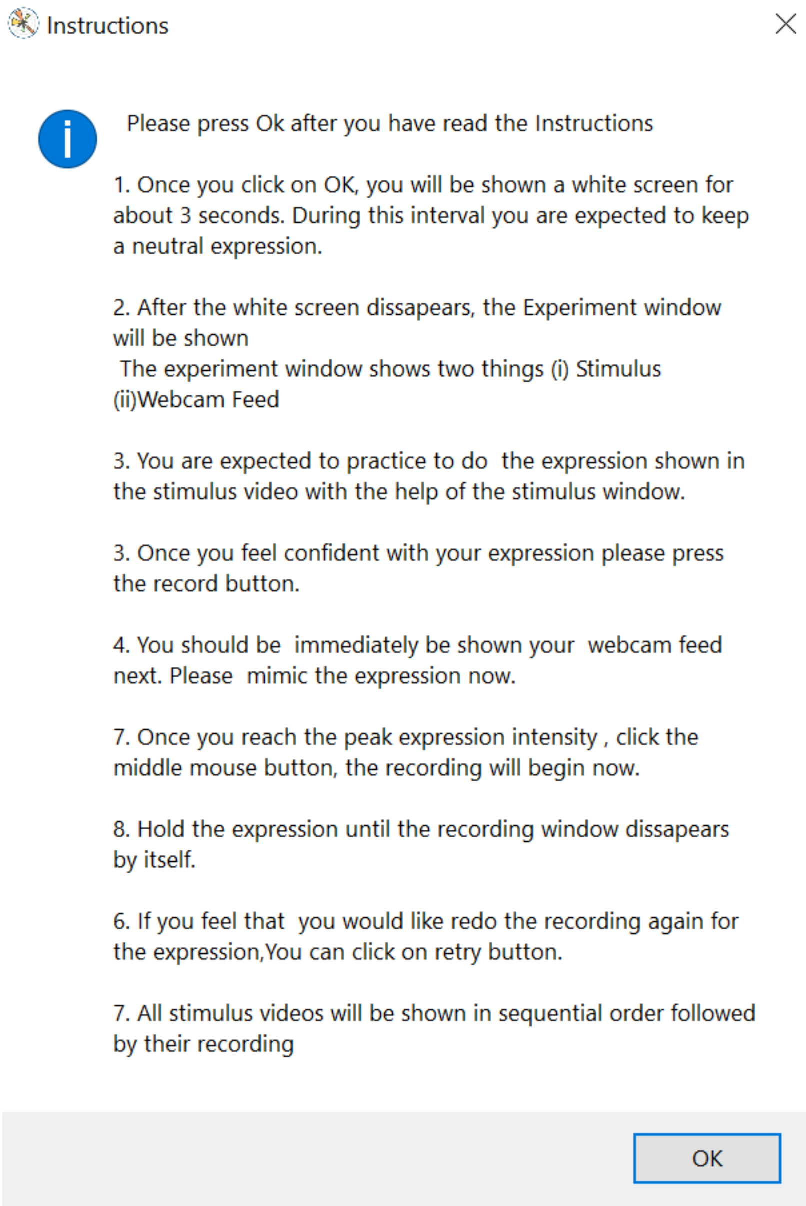

FIGURE 3.11: Window displayed at the start of each trial

to consider the features which can be computed fast. Based on literature material [11][16][1] [21], the following set of features were selected.

3.3.1.1 Statistical Domain

1. **Standard Deviation**

   The difference between each EMG sample and its mean value is represented by the standard

FIGURE 3.12: Experimental Setup

deviation (STD).

$$STD = \sqrt{\frac{1}{N-1}\sum_{1}^{N}(x_i - x_{mean})^2}$$

2. **Variance**

   The Variance of EMG (VAR) expresses the power of the EMG signal as a usable feature.

$$VAR = \frac{1}{N-1}\sum_{i=1}^{N} x_i^2$$

3. **Skewness**

   Skewness is defined as the inclination distribution data. The data is said to have a normal distribution when the location of the average value, the median value and data mode on a line in the curve if these values are not located in one line in the curve occurs the skewness or the heeling

$$skew = \frac{\sum_i^N (x_i - x_{mean})^3}{(N-1) * \sigma^3}$$

4. **Root Mean Square**

   The Root Mean Square (RMS) is modelled as the amplitude modulated Gaussian random process where the RMS is related to the constant force, and the non-fatiguing contractions ofthe muscles

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_i|^2}$$

3.3.1.2 Time Domain

1. **Integrated EMG**

   Integrated EMG (iEMG) is defined as the area under the curve of the rectified EMG signal; that is, the mathematical integral of the absolute value of the raw EMG signal. When the absolute value of the signal is taken, noise will make the mathematical integral have a constant increase.

   $$IEMG = \sum_{i=1}^{N}|x_i|$$

2. **Mean Absolute Value**

   The Mean Absolute Value (MAV) is a technique for assessing and identifying muscular contraction levels. It is calculated by taking the moving average of the full-wave rectified EMG signal.

   $$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i|$$

3. **Simple Square Integral**

   The Simple Square Integral (SSI) expresses the energy of the EMG signal as a usable feature

   $$SSI = \sum_{i=1}^{N}|x_i|^2$$

4. **Waveform Length**

   The Waveform Length (WL) is the total length of the waveform over the segment. The WL calculation's output values represent a measurement of the waveform's amplitude, frequency, and duration.

   $$WL = \sum_{i=1}^{N-1}|x_{i+1} - x_i|$$

5. **Zero Crossings**

   Zero Crossing(ZC) counts the number of times that the sign of the amplitude of the signal changes.

   $$ZC = \sum_{i}^{N} sgn(-x_i x_{i-1})$$

$$sgn(x) = \begin{cases} 1 & if x > 0 \\ 0 & otherwise \end{cases}$$

6. **Number of Peaks**

   Number of peaks is the number of values that are higher than their RMS value.

7. **Mean of Peak Values**

   Mean of peak values is the average of the peak values that have been found in Number of Peaks

8. **Difference Absolute Mean Value**

   Difference Absolute Mean Value (DAMV) is calculated as follows

$$DAMV = \frac{1}{N} \sum_{1}^{N-1} |x_{i+1} - x_i|$$

3.3.1.3 Frequency Domain

1. **Frequency Median**

   The power spectral density (PSD) is used to calculate the Frequency Median (FMD). To determine the frequency domain feature for EMG, there are two ways of PSD estimation: parametric and non parametric. Parametric approaches are based on the assumption that the signal may be treated as the output of a linear system. The non parametric approaches make no assumptions about the system's model. The periodogram method, which was used in our case, is one of the more often used methods.

$$FMD = \frac{1}{2} \sum_{1}^{M} PSD$$

2. **Modified Median Frequency**

   The Modified Median Frequency (MMDF) approach is similar to the FMD method in that it is based on the amplitude spectrum rather than the PSD. It is a frequency expression in which the spectrum is split into two sections of equal amplitude.

$$MMDF = \tfrac{1}{2} \textstyle\sum_{j=11}^{M} A_j$$
$$where\ A_j\ is\ the\ EMG\ Amplitude\ spectrum\ at\ binj$$

3. **Modified Frequency Mean** The Modified Frequency Mean (MMNF) is the average of the frequency based on the amplitude spectrum.

$$MMNF = \frac{\sum_{j=1}^{M} f_j A_j}{\sum_{j=1}^{M} A_j}$$

$$where\ f_j\ is\ the\ frequency\ of\ the\ spectrum\ at\ the\ frequency\ bin\ j$$

Data Preparation

The fEMG data was initially stored in xdf files per recording trial. The frame counts which were also stored in xdf files along with the fEMG data is used to synchronize the data as described in the synchronization section. Following the synchronization procedure, we have the data in the format of fEMG data frames mapped to action units, which are stored in csv files and referred to as pre-processed csv data files. These files can be used as a dataset in future projects.

The pre-processed data is further processed to make it ready for training with the ML models.Data is processed in two different pipelines. The first processing step common to both pipelines is segmentation where we break down the fEMG data into segments of specific lengths. The segments lengths in ms which were considered are 50,100,150,200,250,300,350 and 400. These were taken so as to see the impact of different segment lengths on the model performance. The overview of the processing of pre-processed data is given in Figure 3.13. The final file structure would consist of a data matrix generated for both Raw and Filtered pipelines per segment length.

1. **Raw pipeline**

   This consists of only segmentation followed by feature extraction.

2. **Filter Pipeline**

   This consists of segmentation, band pass filter(15Hz-500Hz),notch filter(50Hz), full wave rectification and finally feature extraction.

3.3.2 ML Model

The problem statement can be described as follows. We are trying to predict the action units based on a frame of EMG data. Since each action unit can be represented as a class. The problem can be reduced to a multi class classification problem.The class labels that are to be predicted are listed in the Table 3.3 and their corresponding actions are listed in Table 3.4. Random Forests

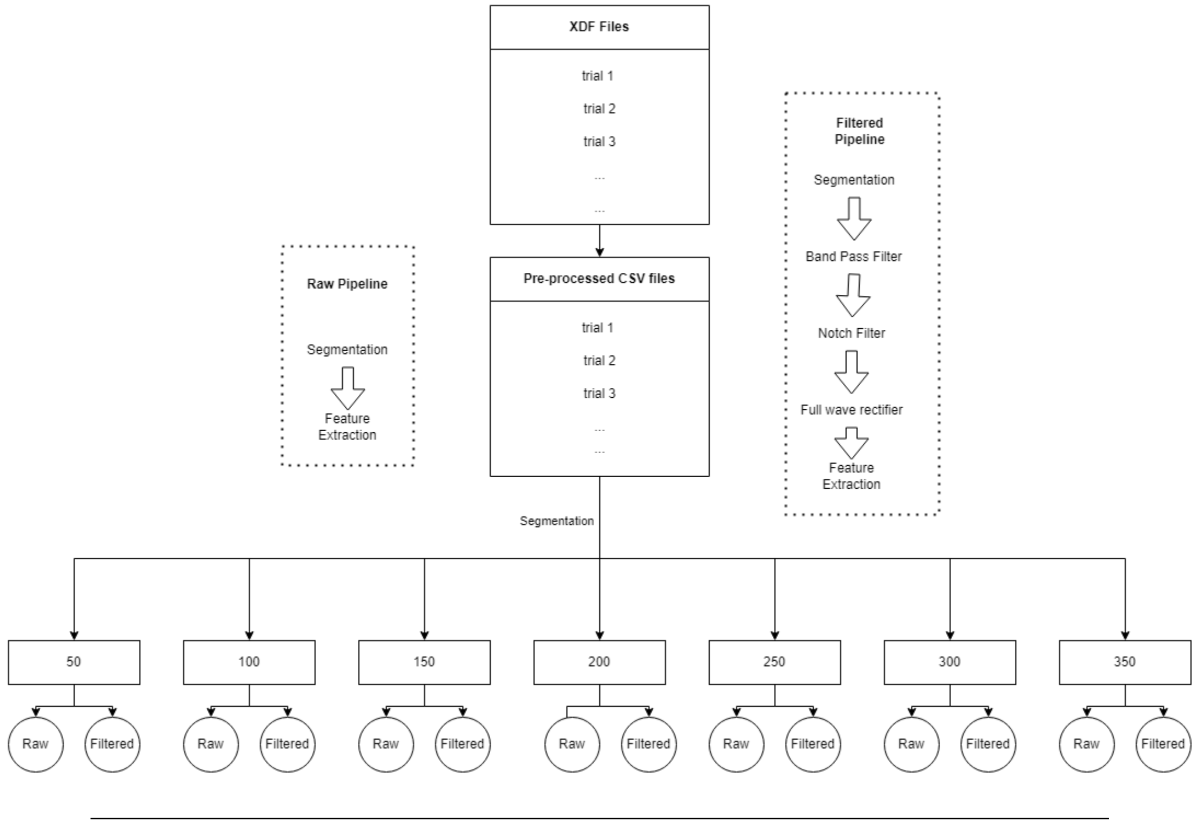

FIGURE 3.13: Data Preparation

was chosen as the initial model of choice as it helps in decoding why a particular class was selected.

Random Forest

The Random Forest module provided by the sklearn API [7] was used for training the model. The training phase consisted of splitting the dataset into two sets i.e training set and testing set using stratified sampling. The training set with the help of the in-built cross validation provided within the randomized grid search modile of the sklearn API was used to train different models with a combination of different parameters until the best model parameters was found. The parameters that were noted to be the best are listed in the table 3.5. These best found parameters were again used to train the models using the training set. The remaining test set was used to evaluate metrics related to the model.

The hardware specification of the host computer on which these models were trained is listed down in Table 3.6

Action Unit Label
AU1
AU2
AU4
AU5
AU6
AU7
AU9
AU10
AU12
AU14
AU15
AU17
AU18
AU20
AU23
AU24
AU25
AU26
AU43

Table 3.3: Action Unit Labels

Action Unit Label	Action
AU1	Inner Brow Raiser
AU2	Outer Brow Raiser
AU4	Brow Lowerer
AU5	Upper Lid Raiser
AU6	Cheek Raiser
AU7	Lid Tightener
AU9	Nose Wrinkler
AU10	Upper Lid Raiser
AU12	Lip Corner Puller
AU14	Dimpler
AU15	Lip Corner Depressor
AU17	Chin Raiser
AU18	Lip Pucker
AU20	Lip Stretcher
AU23	Lip Tightener
AU24	Lip Pressor
AU25	Lips Part
AU26	Jaw Drop
AU43	Eyes Closed

Table 3.4: Action Unit to Action Mapping

Parameter	Value
No. of Estimators	466
Max Depth	50
Max Features	'sqrt'
Bootstrapping	True
Criterion	Gini Impurity

TABLE 3.5: Best Parameters for model

Processor	Intel(R) Core(TM) i7-9700K CPU @ 3.60GHz 3.60 GHz
Installed RAM	16GB
OS	Microsoft Windows 10 Pro

TABLE 3.6: Hardware Specification

Chapter 4

Evaluation

4.1 Task 3: Evaluation

As mentioned in the section related to the data preparation, 8 different segment lengths were considered, which produced 8 different fEMG datasets. Each of these pre-processed datasets was run through two different pipelines , which produced 16 data matrices. For each data matrix, a random forest model with best found hyper parameters was trained. After the models were trained, the accuracy results were obtained on the test sets which was kept aside during the training phase. The accuracy results are showcased in Figure 4.1. The plot demonstrates that, using segments of data of longer lengths are better than shorter segment lengths. One more thing that is noticeable in the plot is, that the data which was not filtered seemed to perform marginally better than the filtered data. This could point to the fact that the frequencies that were filtered out might also be an important asset in action unit recognition.
Further analysis was carried out on the feature importance. The average decrease in Gini impurity was calculated per feature and it was found that the skewness metric calculated from the Corrugator Supercilli channel was found to be the best metric for segregating the samples. These results are showcased in Figure 4.2 . These results were also in agreement with the frequency calculation of the top 5 features across all models in terms of reduction in Gini Impurity 4.3.

The segment length 400 ms was chosen as the basis for further analysis in offline prediction.

.

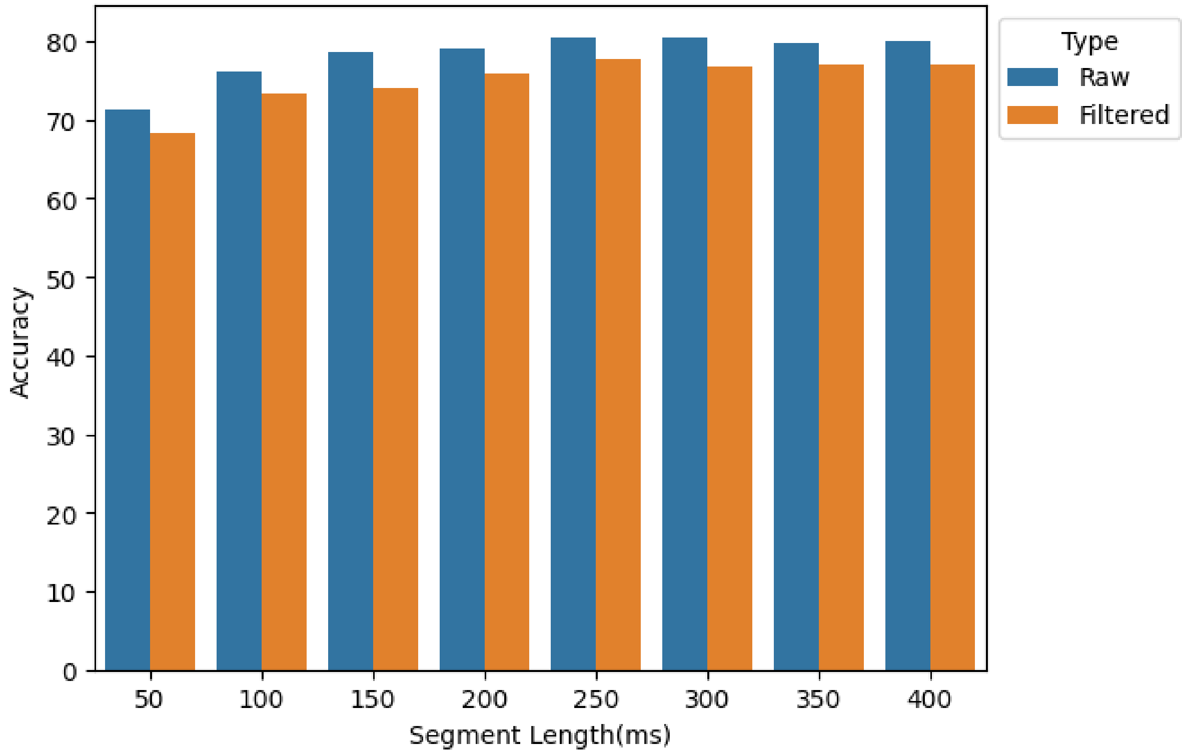

FIGURE 4.1: Accuracy Plot

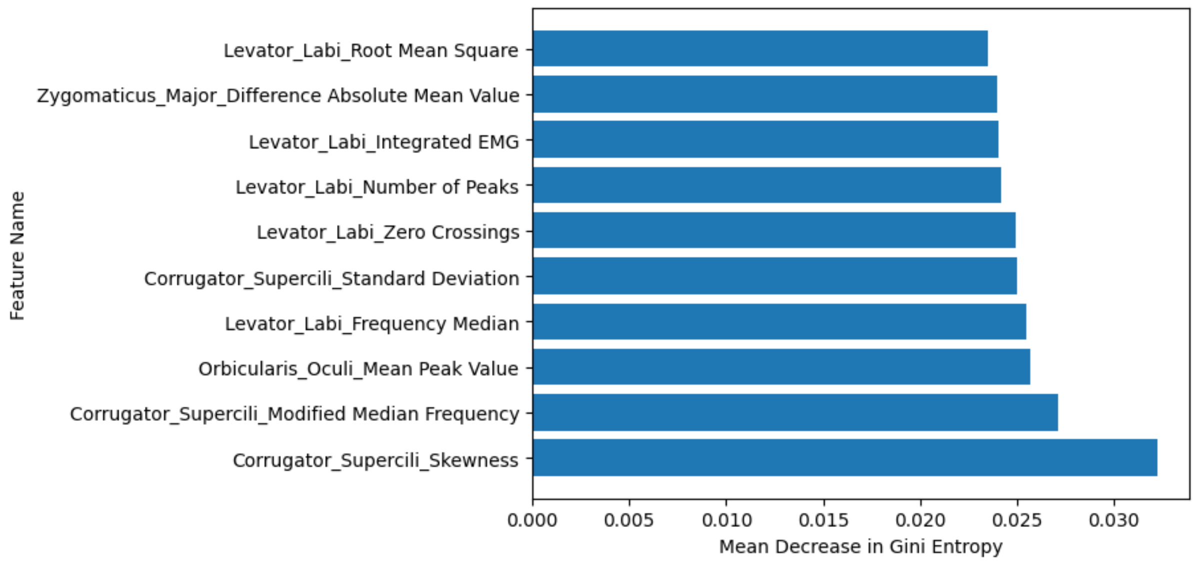

FIGURE 4.2: Avg Decrease in Gini Impurity per feature

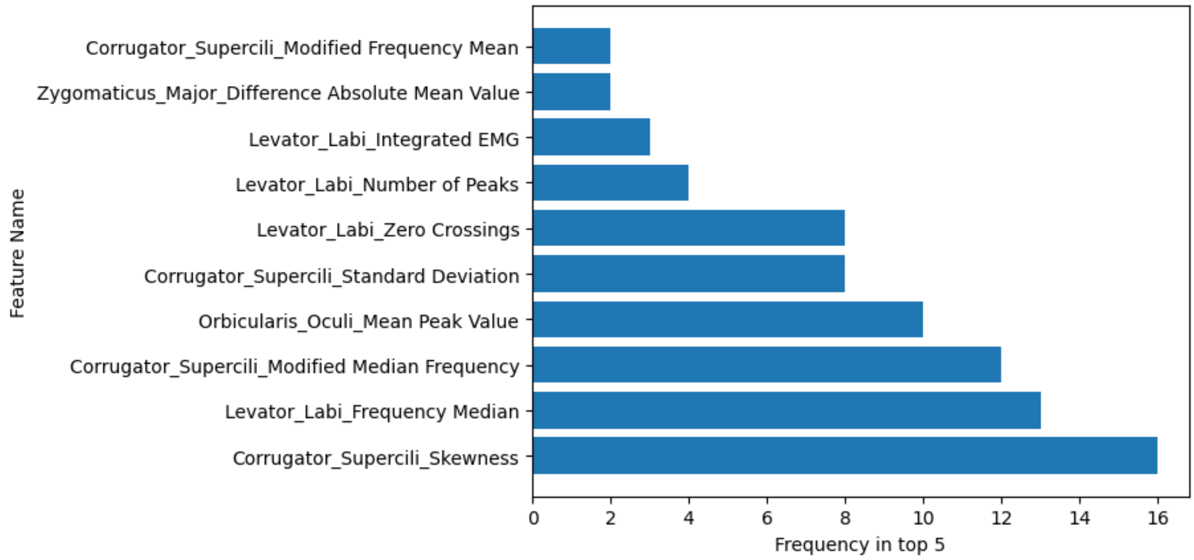

FIGURE 4.3: Frequency Plot

4.2 Offline AU prediction

A recording trial was again carried out which would be used to test our model performance on new data. Classification metrics were calculated out for this set, as this data has not been seen by the model before. First the confusion matrix was calculated for this set which is showcased in Figure 4.4. Apart from this other metrics such as precision ,recall etc. were also calculated which are displayed in Table 4.1. Clearly, the model does not seem to be performing well.However one thing that can be seen from the metrics and the AU predictions that turned to be correct seemed to be concentrated between AU17 & AU23. All of these action units are related to lip moments, as given in Table 3.4. To analyse the this issue more ,instead of predicting the class label , I tried to predict the class probabilities i.e the proportion of decision trees within the that voted for specific classes. These probabilities gave a better sense of how the model was performing. The probabilities output was tested out in real-time to get a more visual sense of what the model was able to identify quickly. How this was done is explained in the next section.

4.3 Output in FACSvatar in real time

A script was created that outputs action units in real time using one of the best performing random forest models(400ms based model). The FACSvatar framework was used to stream the action units into the blender visualization software.Figure 4.5 gives an example of the

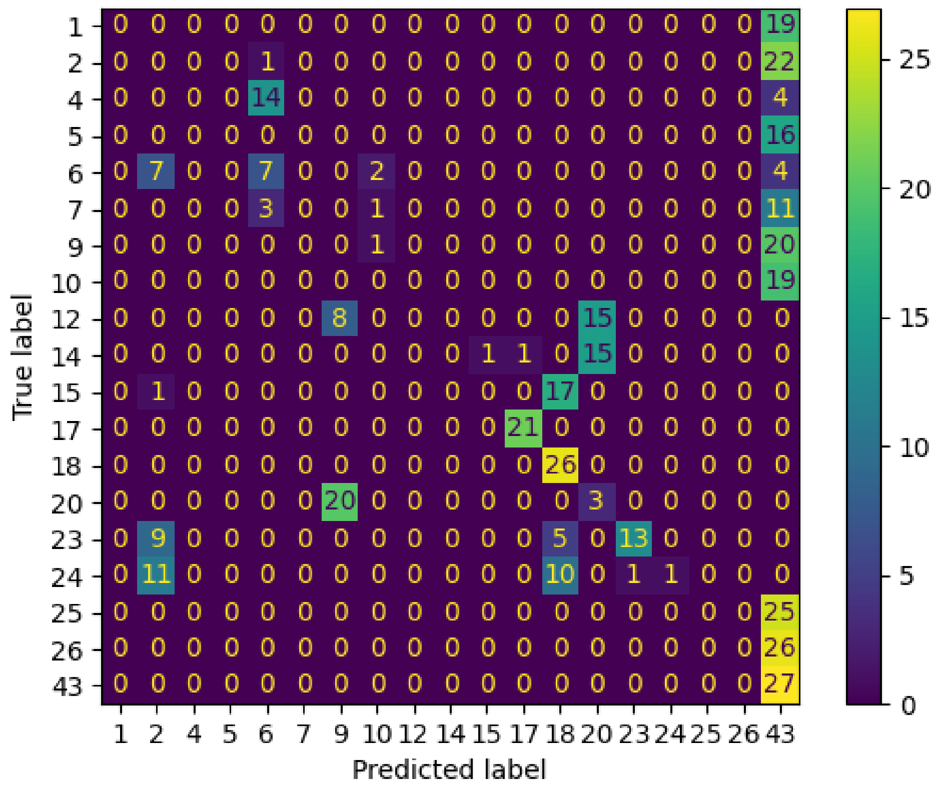

FIGURE 4.4: Confusion Matrix for unseen data. The labels here refer to the action units we are trying to predict

visualization. As the probabilities were themselves low, consequently no discernible output was seen in the visualization software. Despite this, the model does show potential in identifying lip movements. Whenever any expression is made , which involves the movement of the lips. The random forest model seems to be good at picking it up.You can refer to these recording links corresponding to the action units 15,18,20 and 23. Activation of these action units is associated with lip movements as listed in 3.4. Please click on the links to refer to these videos.

4.4 Conclusion

In this Thesis report, I was able to leverage fEMG data to output action unit predictions using a basic Random Forest model. Although the model was failing to predict the correct action unit label in most of the instances. When the probabilities were analysed, some of the decision trees were good at identifying action units associated with lip movements. There is quite a lot of

AU	precision	recall	f1-score	support
1	0.000000	0.000000	0.000000	19.000000
2	0.000000	0.000000	0.000000	23.000000
4	0.000000	0.000000	0.000000	18.000000
5	0.000000	0.000000	0.000000	16.000000
6	0.280000	0.350000	0.311111	20.000000
7	0.000000	0.000000	0.000000	15.000000
9	0.000000	0.000000	0.000000	21.000000
10	0.000000	0.000000	0.000000	19.000000
12	0.000000	0.000000	0.000000	23.000000
14	0.000000	0.000000	0.000000	17.000000
15	0.000000	0.000000	0.000000	18.000000
17	0.954545	1.000000	0.976744	21.000000
18	0.448276	1.000000	0.619048	26.000000
20	0.090909	0.130435	0.107143	23.000000
23	0.928571	0.481481	0.634146	27.000000
24	1.000000	0.043478	0.083333	23.000000
25	0.000000	0.000000	0.000000	25.000000
26	0.000000	0.000000	0.000000	26.000000
43	0.139896	1.000000	0.245455	27.000000
accuracy	0.240786	0.240786	0.240786	0.240786
macro avg	0.202221	0.210810	0.156683	407.000000
weighted avg	0.224177	0.240786	0.174347	407.000000

TABLE 4.1: Classification Metrics

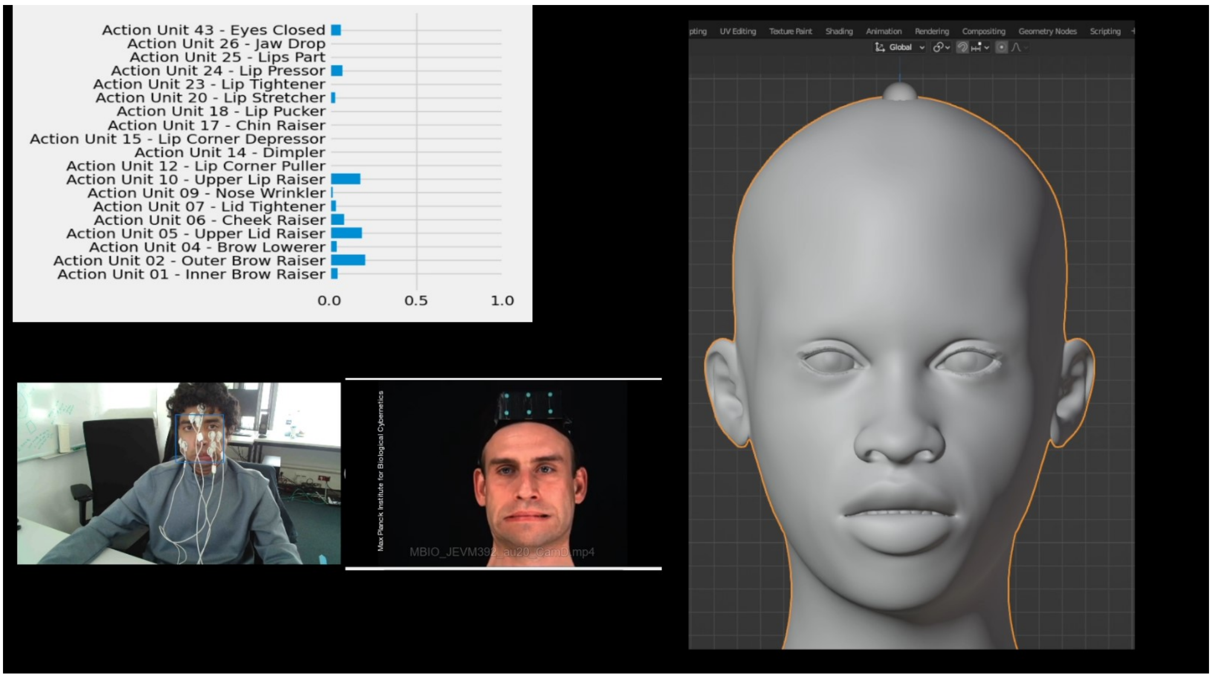

FIGURE 4.5: Real Time AU visualization

potential avenues that can be explored with the help of the findings. Some of the options that will be explored in the future are

1. More participant recordings to help gauge the model performance metrics, using Leave One Subject validation.
2. More features to be tried out. TSFEL(Time Series Feature Extraction Library) [3] library will be use out to try all kind of features suitable with EMG data.
3. Try out more ML models like Neural Networks and Support Vector Machines, to see if they make any difference.
4. Multi label classification instead of Multi class classification as performing any facial expression includes the activation of several action units instead of only one.

Appendix A

Comparison of AU recognition software -In the following link, you will find the comparison videos used to judge the capabilities of each AU recognition software.

Stimulus Videos -In the following link, you will find the stimulus videos used during the experimental trials.

Comparison of AU recognition software with Blender -In the following link, you will find the comparison videos used to judge the capabilities of each AU recognition software in blender.

Comparison of AU output by ML model in real time -In the following link, you will find the comparison videos used to judge the capabilities of ml model output in real time.

Github Repository -In the following link, you will source code for the thesis.

Bibliography

[1] Sara Abbaspour et al. "Evaluation of surface EMG-based recognition algorithms for decoding hand movements". In: *Medical &amp Biological Engineering &amp Computing* 58.1 (Nov. 2019), pp. 83–100. DOI: 10.1007/s11517-019-02073-z. URL: https://doi.org/10.1007/s11517-019-02073-z.

[2] Tadas Baltrusaitis et al. "OpenFace 2.0: Facial behavior analysis toolkit". In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. Xi'an: IEEE, May 2018.

[3] Marília Barandas et al. "TSFEL: Time Series Feature Extraction Library". In: *SoftwareX* 11 (Jan. 2020), p. 100456. DOI: 10.1016/j.softx.2020.100456. URL: https://doi.org/10.1016/j.softx.2020.100456.

[4] Lisa Feldman Barrett et al. "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements". In: *Psychological Science in the Public Interest* 20.1 (2019). PMID: 31313636, pp. 1–68. DOI: 10.1177/1529100619832930. eprint: https://doi.org/10.1177/1529100619832930. URL: https://doi.org/10.1177/1529100619832930.

[5] A. van Boxtel. "Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles". In: *Psychophysiology* 38.1 (2001), pp. 22–34. DOI: https://doi.org/10.1111/1469-8986.3810022. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1469-8986.3810022. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1469-8986.3810022.

[6] Leyde Briceno and Gunther Paul. "MakeHuman: A Review of the Modelling Framework". In: *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. Ed. by Sebastiano Bagnara et al. Cham: Springer International Publishing, 2019, pp. 224–232. ISBN: 978-3-319-96077-7.

[7] Lars Buitinck et al. "API design for machine learning software: experiences from the scikit-learn project". In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. 2013, pp. 108–122.

[8] Paul Ekman and Wallace V Friesen. "Facial action coding system". In: *Environmental Psychology Nonverbal Behavior* (1978).

[9] Alan J. Fridlund and John T. Cacioppo. "Guidelines for Human Electromyographic Research". In: *Psychophysiology* 23.5 (1986), pp. 567–589. DOI: `https://doi.org/10.1111/j.1469-8986.1986.tb00676.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.1986.tb00676.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8986.1986.tb00676.x`.

[10] Michaël Gilbert, Samuel Demarchi, and Isabel Urdapilleta. "FACSHuman, a software program for creating experimental material by modeling 3D facial expressions". en. In: *Behav. Res. Methods* 53.5 (Oct. 2021), pp. 2252–2272.

[11] Mahyar Hamedi et al. "Comparison of Different Time-Domain Feature Extraction Methods on Facial Gestures? EMGs". In: vol. 12. Apr. 2012.

[12] Xiaohua Huang et al. "Towards a dynamic expression recognition system under facial occlusion". In: *Pattern Recognition Letters* 33 (Dec. 2012), pp. 2181–2191. DOI: `10.1016/j.patrec.2012.07.015`.

[13] Mario Kleiner et al. "Multi-viewpoint video capture for facial perception research". In: *Captech 2004 - Workshop on modelling and motion capture techniques for virtual environments, 55-60 (2004)* (Jan. 2004).

[14] Christos Kyrlitsias and Despina Michael-Grigoriou. "Social Interaction With Agents and Avatars in Immersive Virtual Environments: A Survey". In: *Frontiers in Virtual Reality* 2 (2022). ISSN: 2673-4192. DOI: `10.3389/frvir.2021.786665`. URL: `https://www.frontiersin.org/articles/10.3389/frvir.2021.786665`.

[15] Gwen Littlewort et al. "The computer expression recognition toolbox (CERT)". In: *Face and Gesture 2011*. IEEE, Mar. 2011. DOI: `10.1109/fg.2011.5771414`. URL: `https://doi.org/10.1109/fg.2011.5771414`.

[16] Radek Martinek et al. "Advanced Bioelectrical Signal Processing Methods: Past, Present, and Future Approach—Part III: Other Biosignals". In: *Sensors* 21.18 (2021). ISSN: 1424-8220. DOI: `10.3390/s21186064`. URL: `https://www.mdpi.com/1424-8220/21/18/6064`.

[17] Mark Nestor, Glynis Ablon, and Andy Pickett. “Key Parameters for the Use of AbobotulinumtoxinA in Aesthetics: Onset and Duration”. In: *Aesthetic Surgery Journal* 37.suppl_1 (Apr. 2017), S20–S31. DOI: 10.1093/asj/sjw282. URL: https://doi.org/10.1093/asj/sjw282.

[18] Melanio Noe O. Manipon and https://orcid.org/0000-0003-0104-5465, neorico27.mnm@gmail.com, Saint Mary’s University, Philippines. “Effectiveness of ChemiCooking as A gamified intervention in nomenclature of compounds: Learning experiences of Grade 11 students in A public school”. In: *International Multidisciplinary Research Journal* 5.1 (Feb. 2023).

[19] Melanio Noe O. Manipon and https://orcid.org/0000-0003-0104-5465, neorico27.mnm@gmail.com, Saint Mary’s University, Philippines. “Effectiveness of ChemiCooking as A gamified intervention in nomenclature of compounds: Learning experiences of Grade 11 students in A public school”. In: *International Multidisciplinary Research Journal* 5.1 (Feb. 2023).

[20] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin. “Techniques of EMG signal analysis: detection, processing, classification and applications”. In: *Biological Procedures Online* 8.1 (Dec. 2006), pp. 11–35. DOI: 10.1251/bpo115. URL: https://doi.org/10.1251/bpo115.

[21] Christopher Spiewak et al. “A comprehensive study on EMG feature extraction and classifiers”. In: *Open Access Journal of Biomedical Engineering and Biosciences* 1.1 (2018), pp. 1–10.

[22] Stef van der Struijk et al. “FACSvatar: An Open Source Modular Framework for Real-Time FACS Based Facial Animation”. In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. IVA ’18. Sydney, NSW, Australia: Association for Computing Machinery, 2018, 159–164. ISBN: 9781450360135. DOI: 10.1145/3267851.3267918. URL: https://doi.org/10.1145/3267851.3267918.

[23] Yuanjie Wu et al. “Using a Fully Expressive Avatar to Collaborate in Virtual Reality: Evaluation of Task Performance, Presence, and Attraction”. In: *Frontiers in Virtual Reality* 2 (2021). ISSN: 2673-4192. DOI: 10.3389/frvir.2021.641296. URL: https://www.frontiersin.org/articles/10.3389/frvir.2021.641296.