Abstract

Sentiments and emotional analysis have recently come up to become a very prominent topic in various research and programming fields. From customer review to analyzing public's opinion, various visualization techniques have evolved over recent years on such attributes which involve convoluted datasets and bringing out relations out of them. Various approaches have been identified in analyzing sentiments and emotions like reading pupil diameter, vocal-sound signals and text analysis. This report introduces the audio-visual approach, in which emotions of participants are captured, analyzed and compared with the participant's self-evaluation. The emotions are converted to sentiments in order to determine the correlativity between the implemented system and self-evaluation. The results procured are later analyzed on basis of relational and geospatial aspects. Further, lexical approach is used in identifying which particular lexicon is accurate and efficient enough in determining the accurate sentimental valiance. This survey is useful for researchers whose interests lie in sentiments-emotional analysis and in the visualization techniques.

Table of Contents

List of Abbreviations

HCI	**:**	Human Computer Interactions
SO-CAL	**:**	Semantic Orientation CALculator
CNS	**:**	Central Nervous System
CSD	**:**	Chrome Shape Detection
API	**:**	Application Programming Interface
CNN	**:**	Convolutional Neural Networks
FER+	**:**	Facial Expression Recognition
RAF-DB	**:**	Real-world Affective Faces Database
PPMC	**:**	Pearson Product Moment Correlation

List of Figures

List of Tables

1. Introduction

Human’s explicit their feelings in various different ways, it can be vocal, facial or through signs or gestures. It is within the normal understanding of a human to identify what the other person is trying to communicate through such mutual understanding of verbal and non-verbal expressions. Even with difference in languages, cultures and ecological background there are certain characteristic’s which defines every human and can be considered as a common trait among us all, they are emotions and sentiments. It was the naturalist Charles Darwin who proposed the Theory of Emotions: ‘It is through emotions that we humans and animals have survived throughout this time’ [1]. This simply proof’s that emotions do have a very significant impact in our perception, decision making, followed by our action generation and control on the consequences of our actions. When considering the technical aspects about how human interactions with computers too have evolved, it raises a peculiar question; can a computer or a system too understand the emotions and sentiments of a human? Can computer too identify the emotional and sentimental aspects of a human irrespective of ecological and cultural background? With today’s globalization, recognition of emotional and sentimental expressions is of key importance for development of HCI systems. Taking into account the amount of complexities in sentiment and emotional recognition, much research has been conducted to understand the mechanism involved which is driven by the huge amount of promising usages and benefits such systems can have in future. This project is the first stepping stone for the group, which involves implementation of a system that captures emotions in real time and stores them locally in a form of text file. The emotions and sentiments have later been analysed and visualized to share out the results obtained throughout the experiment. In field of social and clinical psychology, emotion and sentiment detection systems could help to diagnose psychological disorders including other side effects like fating, stress, and depression at their very early stages. Organizations and institutions can make use of such systems in identifying emotional symptoms (like anger, fear, sad, shock) which form basis for emotional exhaustion.

1.1 Emotional-Sentiment Analysis

Very often terms like emotions and sentiments are used interchangeably. But, are they actually the same concept? It does make sense after all sentiment is a kind of feeling or emotion and the two forms of analytics are often closely related to each other [2]. By dividing emotions between two polarities positive and negative a lot can be perceived and obtained. In the later part of this report the emotions are later converted to sentiments in order to determine the correlativity between the system and participants.

1.2 Scope and Objective

In this research, one first needs to understand that human psychology is a very complex process and integrating this domain with the technical aspect ultimately brings in more interesting challenges. During the course of this experiment, various reactions of the participants were observed and analyzed which allowed drawing out some hypothetical questions:

1. What is the correlativity of sentiments between the system and participants?
2. Do emotions/sentiments differ or are similar among humans (participants) irrespective of their ecological and cultural background?
3. Do participants relate themselves with the situation presented to them in the interview, like has it ever happened to them in their life?
4. Which lexical approach would be better if in-case the system is fully automated in future?

Other than the above mentioned hypothesis, few other observations were analysed which are briefly illustrated throughout this report.

2. Related Work

There are various studies in the field of emotional and sentiment analysis, most of these have been categorised on the basis of resources and methods used. Taboada, Brooke, & Stede presented a lexicon based approach to extract sentiments from texts. Their lexicon is a dictionary of words which are categorised into two polarities; positive and negative. By using SO-CAL, they have created their own dictionary of words on basis of which they have concluded the lexicon based approach which is robust and results in good cross-domain performance. [3]

Kucher, Paradis & Kerren have described '*The State of Art in Sentiment Visualization*', in which they have introduced various sentiment visualization techniques from text data and have categorised them into 7 groups with 35 sub-categories. Overall they have discussed about 132 visualization techniques originating from peer viewed publications. [4]

The focus of sentiment and emotional analysis research has shifted towards social media, primarily targeting Twitter, Facebook, and YouTube. Analysing such platforms has benefited various organizations in order to understand the emotions and opinions of public. The study by Ohman, Honkela & Tiedemann, has introduced a lexical approach on sentimental analysis across multiple languages (Spanish, Portuguese, Finnish and English). Their research shows that study of multilingual sentiments from social media platform is important as it grants the possibility to compare how people from different ecological and cultural background view various topics. [5]

Another study by Sauter, et al., has examined emotions across two groups from different cultural background using vocal signals. Their finding show's that number of negative emotions have vocalizations that are recognizable across cultures, while positive emotions are recognized cultural specifically. Further, basic emotions (anger, fear, disgust, joy, sad and surprise) were identifiable by both the groups which support the theory that these emotions are psychologically universal and are shared by all humans. [6]

3. High Level System Architecture

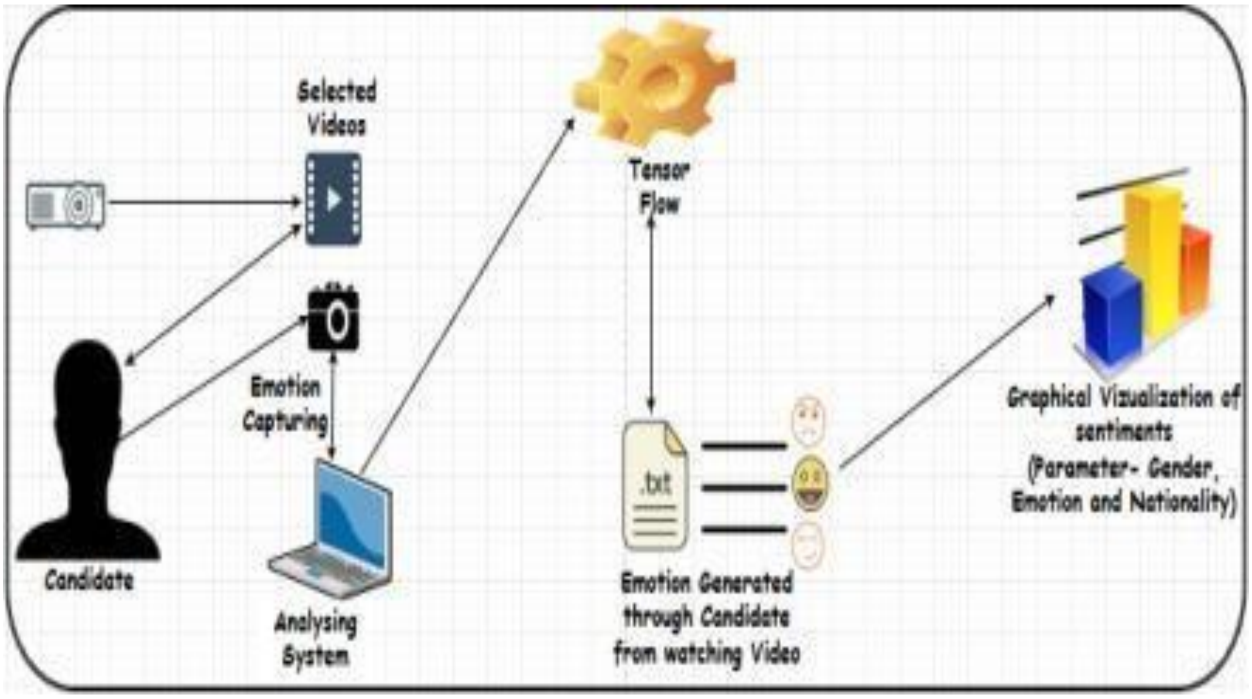


Fig.1: System Architecture

The above architecture shows working of the proposed model from initial to final stage of the process. In the first phase, a participant is seated in a favourable environment to induce effective stimulation via the selected videos. The favourable environment is explained in later parts of this report. The implemented system is placed in-front of the participant in close range such that the system can easily capture emotions without any hindrance for further analysis. The selected videos are played through projector with a great impact of speakers to give a real feel of the situation, which assisted in increasing the quality of stimulation being delivered.

With integration of **Tenserflow.js** along with facial detection model **face-api.js**, the emotions are captured in real time simultaneously when the participant is shown one of the selected videos. The emotions are being collected from a laptop's webcam on which the proposed model is implemented. On the basis of facial reaction made by participants, the system highlights and predicts the type of emotion being captured and stores the captured emotions in a digital text format (.txt file). For security and privacy concerns it was decided to design system such that it only predicts and stores emotions in text format rather than recording the participant's facial reaction.

The recorded text file contains relative information of the participant like first and last name followed by gender, age and nationality which a participant has to manually key-in before starting the system. Along with these relative details, the text file also holds emotions (happy, surprise, neutral, fear, sad, angry and disgust) which are predicted and captured by the system. Further, to compare it with real life experience of candidate, a self-evaluation form as per the selected question was filled in by every interviewed participant. This gave support to this experiment of sentimental analysis about the accuracy of implemented system by correlating the result of system with manual answer of the candidate.

In the final stage of this project, two forms of datasets were collected. The collected datasets were analysed with relational and geospatial aspects and visualized graphically. Later, the same dataset were analysed using the lexical and text-mining approach in R-studios to understand the emotional and sentimental tone of participants. The lexical approach was used in order to identify which lexicon is accurate if in-case system were supposed to be fully automated in future. Through this approach emotional valance of the participants can be observed.

4. Methods & Procedure

4.1 Implementation

The objective of the implementation part was to design and create a system which could accurately capture emotions of an individual. After a lot of research a system was successfully implemented which capture's emotions in real time. The functioning of this system along with the methods integrated is explained below.

4.1.1 Data Sets

The datasets used in this project are **FER+** and **RAF-DB**. **FER+** which comprises of 48x48 pixels of grayscale facial pictures. These facial pictures have been enlisted such that the confront is more or less centred and involves relatively the same sum of space in each picture. Each confront is categorised into seven emotions **(0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral)** on the basis of facial expression. This comprised of 28,709 such examples out of which 3,589 were determined as the actual test set.

Fig.2: Examples from FER+ Database

Since, the system was also to be tested on participants from diverse and cross-cultural background, **RAF-DB** which is a database for facial expression with about Thirty-Thousand diverse facial expressions was used too. **RAF-DB** is based upon the annotation of crowdsourcing where each image is individually labelled by forty annotators. This database comes with great set of variables like ethnicity, age and gender etc. There are other important features that **RAF-DB** consists of: real-world images; baseline output for emotions; Five-facial landmark location (37 landmark location, race, age, gender, box bounding); distribution of emotions in vector format.

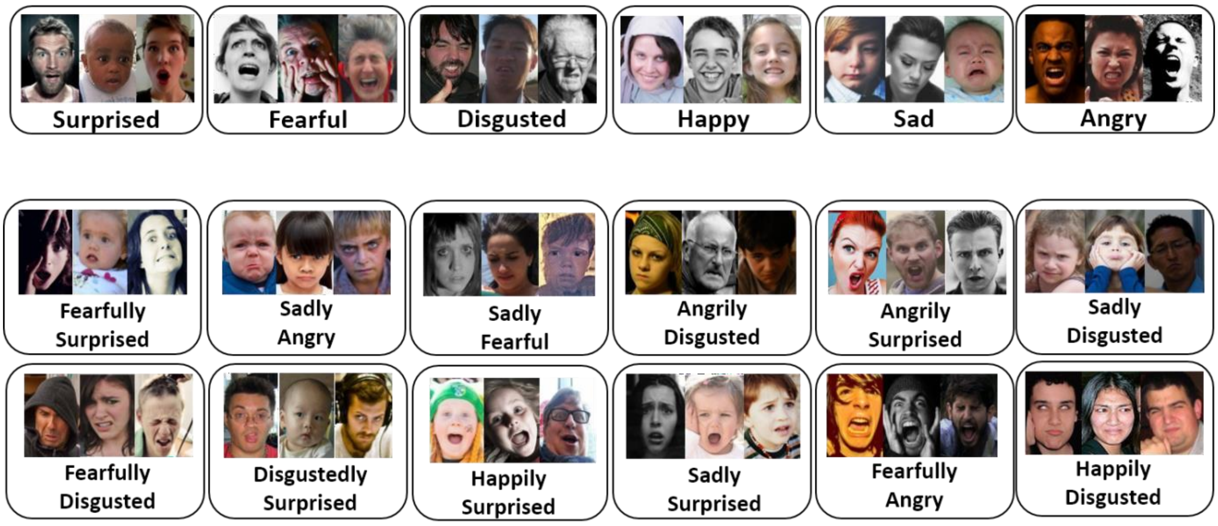


Fig.3: Examples from RAF-DB

The emotions labelled in the dataset from **RAF-DB:**

0: -4593 images- **Angry**

1: -547 images- **Disgust**

2: -5121 images- **Fear**

3: -8989 images- **Happy**

4: -6077 images- **Sad**

5: -4002 images- **Surprise**

6: -6198 images- **Neutral**

4.1.2 Facial Detection

For facial detection two integrated methods were used **Chrome Shape Detection API** and **face-api.js**. In CSD-API, the images and photos used were mostly composed from the internet holding identifiable liniments like human faces and data in text format. Identifying these kinds of features are costly when done computationally. Through this API, computing hardware like webcam and other camera based devices can detect human faces.

Further through this API certain functions like **FaceDetector(optional FaceDetectorOptions faceDetectorOptions)** which constructs a new FaceDetector whenever the application runs, **detect(ImageBitmapSource image)** which detects the human faces, **maxDetectedFaces** which allows limiting the number of facial detection are integrated.

The **face-api.js** is a JavaScript module which was built over **tenserflow.js core.** These implement multiple CNN's to apply and solve facial detection, recognition and facial landmarks over the web-portal. Through this, the implemented system can be optimized to other portable devices like tablets and mobile devices thus making the system platform independent. The **face-api.js** was majorly used as it provided models which allowed implementing facial recognition over the browser.

4.1.3 Tensorflow.js

Tenserflow.js is an open source JavaScript based library which can be used in implementing a pre-trained model or modify the pre-trained model as per desired requirement. In this project **Tenserflow.js,** was used as a back-end which held various pre-trained models from mentioned datasets over which the **face-api.js** was integrated upon. By integrating various CNN's through **face-api.js**, the face recognition model was successfully implemented over browser without using any external dependencies.

4.1.4 Emotion Detection

For emotion detections the **TinyFaceDetector** model was used which along with **face-api.js** was able to predict the emotions by the facial reactions in real-time. The **TinyFaceDetector** is a pre-trained model which consists of about fourteen thousand labelled images with bounding boxes. The **TinyFaceDetector** along with **face-api.js** implements face landmark detector which detects 68 facial points, this (**faceLandmark68Net**) model's actual size is 350KB and is trained with thirty-five thousand labelled images each marked with 68 facial points. Further, **face-api.js** enabled another method called **faceRecognitionNet** which allows the API to recognize whether the face is in the box or not and in the end an essential method **faceExpressionNet** to detect the types of emotions on the basis of **faceRecognitionNet.**

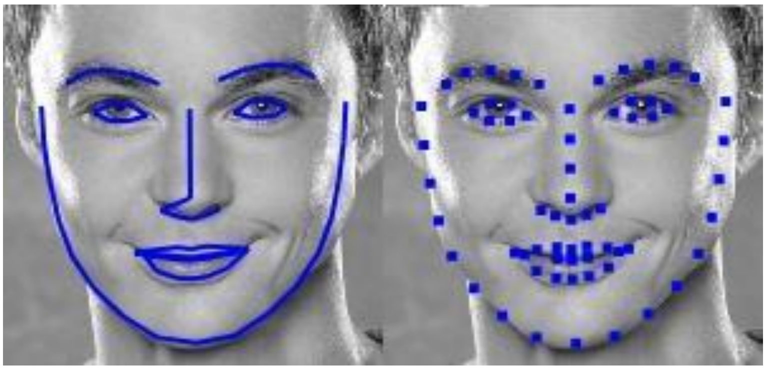


Fig.4: Example for Facial-Points

4.2 System Work-flow

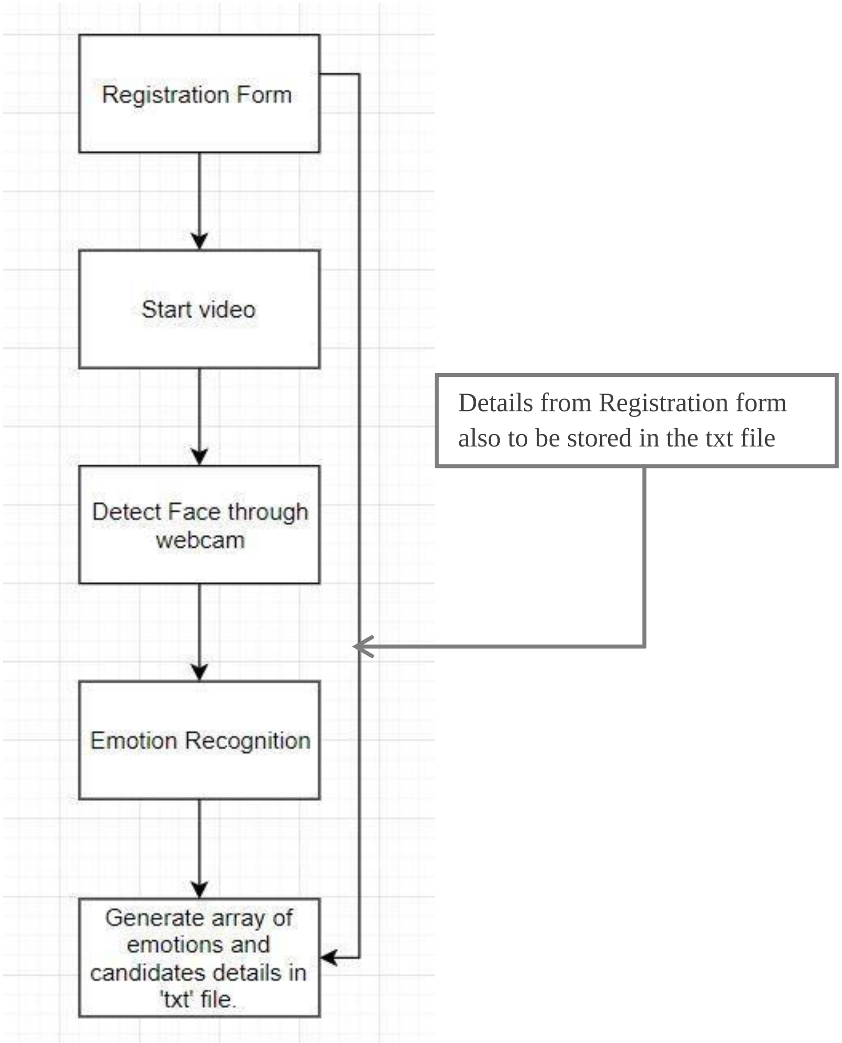


Fig.5: Emotion-Detection System Workflow

The above illustrated flow chart explains the functionality of the implemented emotion detection system. The participant enters their basic information like first and last name, age, nationality, gender. Once the information is submitted, the system will initiate the webcam, simultaneously video from another system is manually started to induce emotions. The system will begin detecting the emotions on basis of participant's facial expressions. Once the video is completed the emotion detection system is stopped by clicking on the submit button, which will further generate a text file which will hold the submitted information from participant along with the array of emotions detected by the system.

4.3 Detection Results

This is the final phase of how exactly the system would appear when detecting emotions out of participants facial expressions. For achieving this part a canvas was created with a rectangular box with different colour codes along with text. In text, one array of each seven emotions (angry, disgust, fear, happy, sad, surprise, neutral) was defined. Every rectangular box frame for every emotion were colour coded (“#ff0000“(red), "#00a800“(green), "#ff4fc1“(pink), "#ffe100“(yellow), "#306eff“(blue), "#ff9d00“(orange), "#7c7c7c”(gray)). Further, to create a web-page for ‘Registration form’ which would be store relative data along the with emotions in text file, following methods were used.

DrawBox drawing options: - export interface IDrawBoxOptions { boxColor?: string lineWidth?: number drawLabelOptions?: IDrawTextFieldOptions label?: string }

TextFieldOptions : - backgroundColor?: string fontColor?: string fontSize?: number fontStyle?: string padding?: number

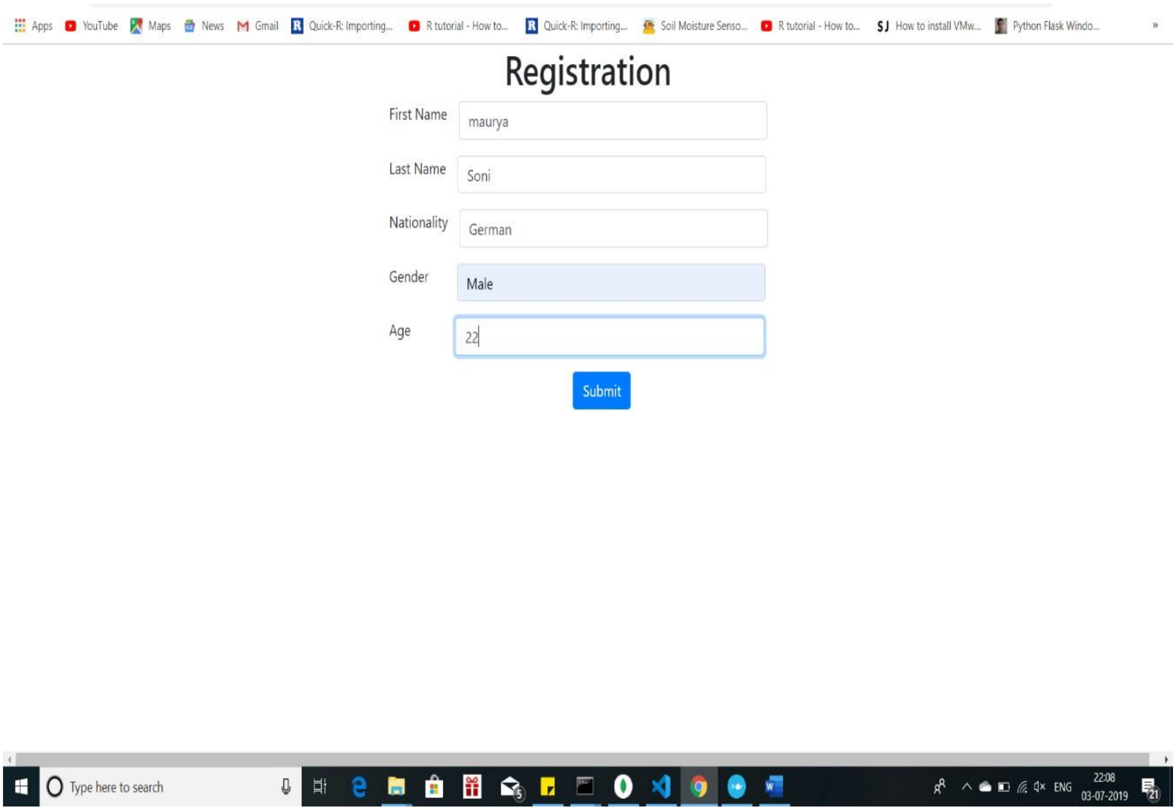


Fig.6: Registration Form

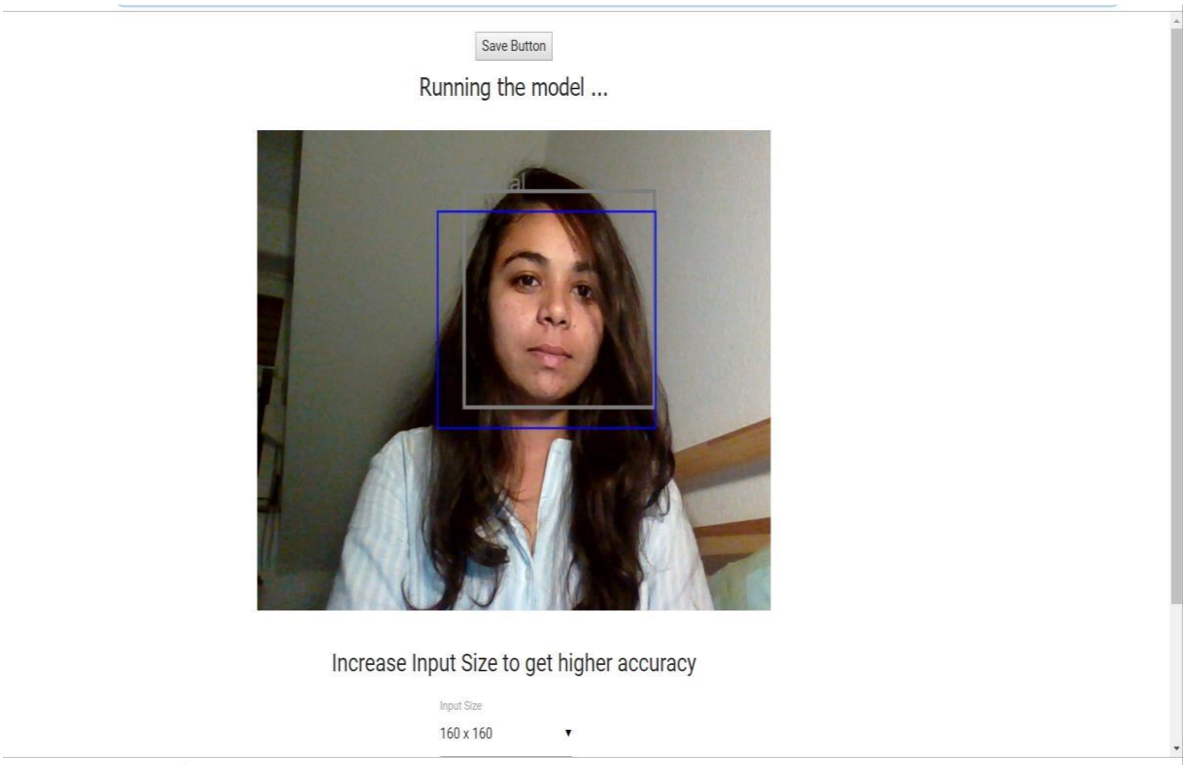


Fig.7: System Capturing Emotions in Real Time

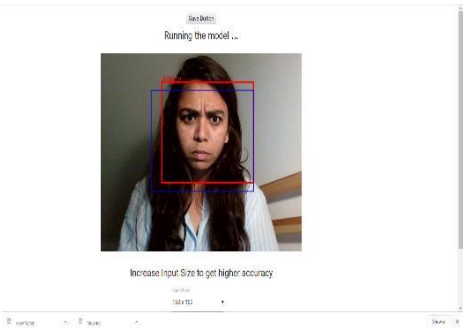


Fig.8: Angry (Red)

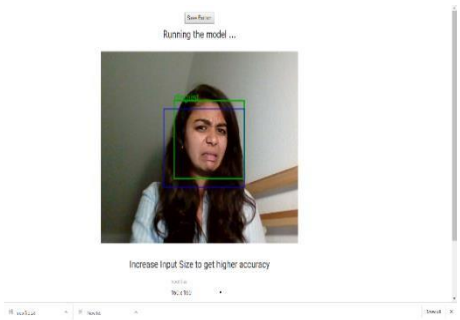


Fig.9: Disgust (Green)

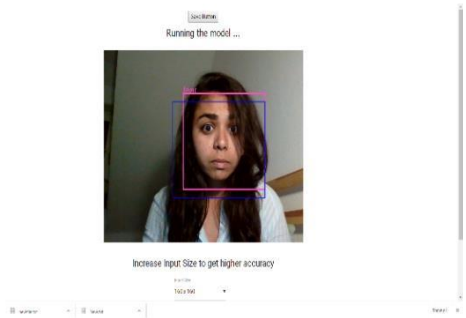


Fig.10: Fear (Pink)

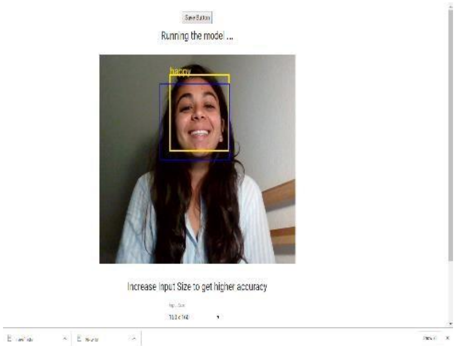


Fig.11: Happy (Yellow)

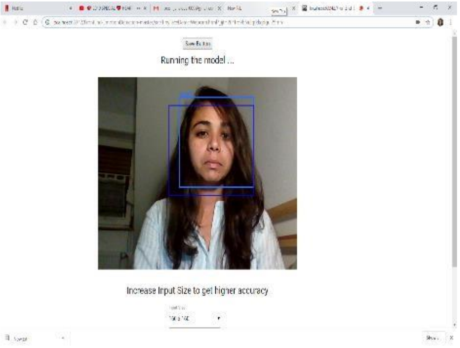


Fig.12: Sad (Blue)

Fig.13: Neutral (Grey)

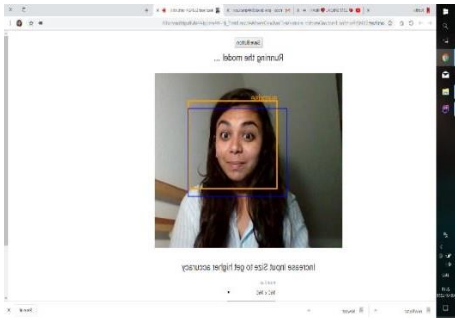


Fig.14: Surprise (Orange)

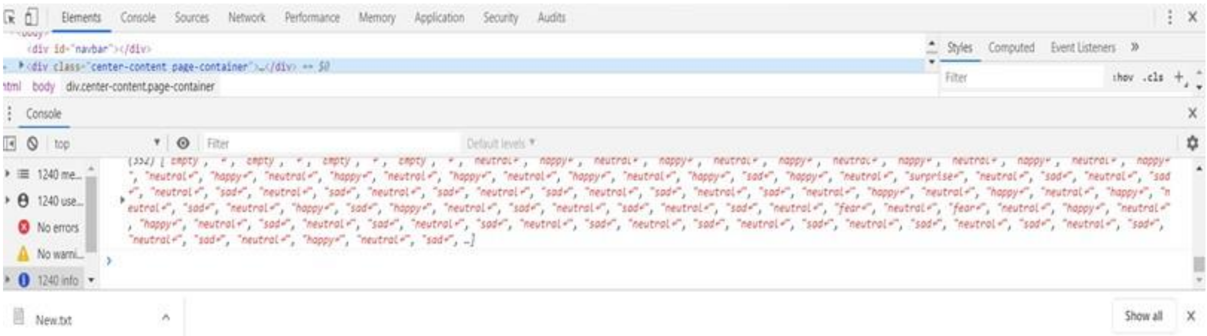


Fig.15: Array of Emotions

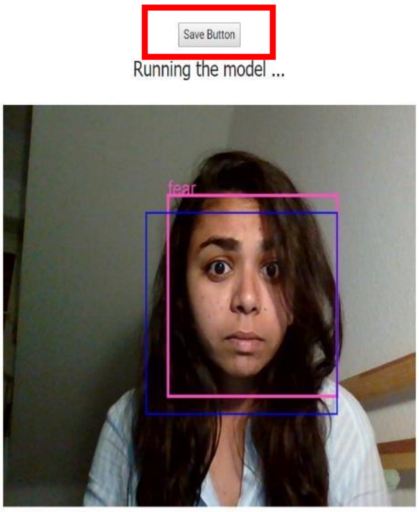


Fig.16: Click on save button to stop model from detecting emotions

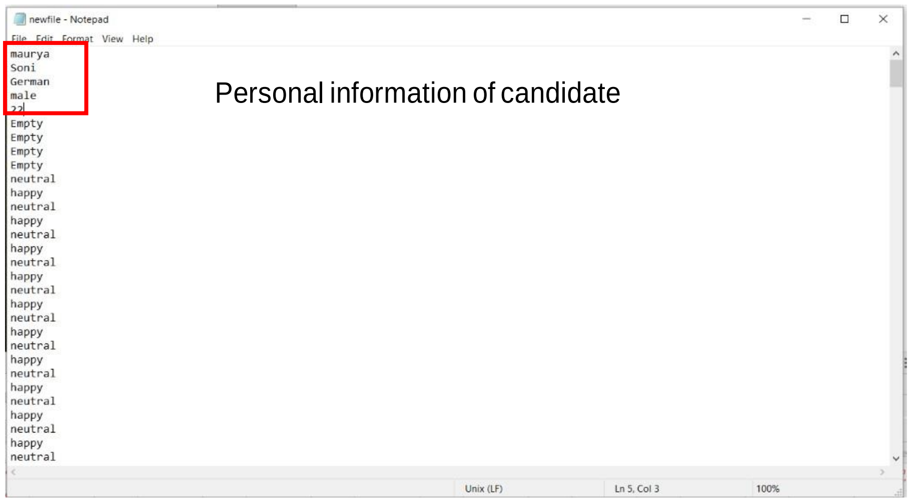


Fig.17: System downloads .txt file with personal information and emotions

4.4 Participants

The targeted participants for this project were students of the Rhein-Waal University, Kamp-Lintfort. The mean age of the participants interviewed was Twenty-four (24). Selecting University students for interviewing provided a cross cultural advantage which was an important aspect for this project. About forty-nine participants (Female: 26 & Male: 23) with normal vision participated in the experiment. The participants were not afflicted by any kind of medication to influence any sort of emotions. The majority of the participants were native Germans followed by Indians. Briefings about the experiment were given to every participant and were also informed that no personal information other than the nationality of the participant would be revealed.

4.5 Stimuli

Stimulation in general is used to trigger emotions and in such type of experiments can be categorized on basis of methods and procedures used. In this experiment the concept of audio-visual approach was used which is based upon the incitement of the CNS. Such sort of stimulation effects the brain in various ways depending upon the visual and sound effects delivered.

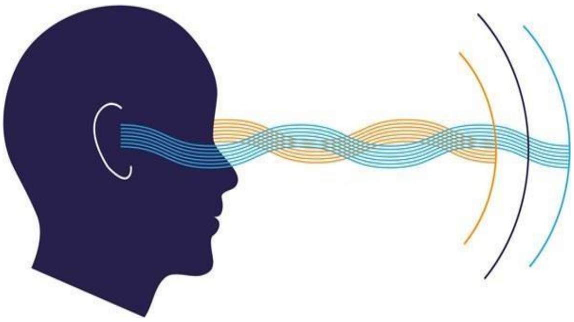


Fig.18: Example Flow of Audio-Visual Stimulation

To ensure effective and spontaneous occurrence of emotions, a strong stimulation was indeed required, which lead to another challenge of selecting videos which could deliver strong stimuli to the participants. Hence, it was decided to select videos which follow the basic ethics of psychology and should be shorter in length to consume less of participant’s valuable time. Also, environmental factors or surroundings also do play an important role in delivering effective stimulation; all these factors are discussed in following section of this report. [7]

4.5.1 Videos Selection

This was another important phase for this project. Since there were no earlier data available for audio-visual effects which could be used in stimulating emotions, it was decided to select set of videos which could influence some emotions in the participants. Five different videos were selected keeping in mind the following criteria:

1. Each video should be able to influence certain type of emotion
2. No political and cultural influence
3. Easy to understand and perceive irrespective of language
4. Short in length
5. Follow the ethics of psychology

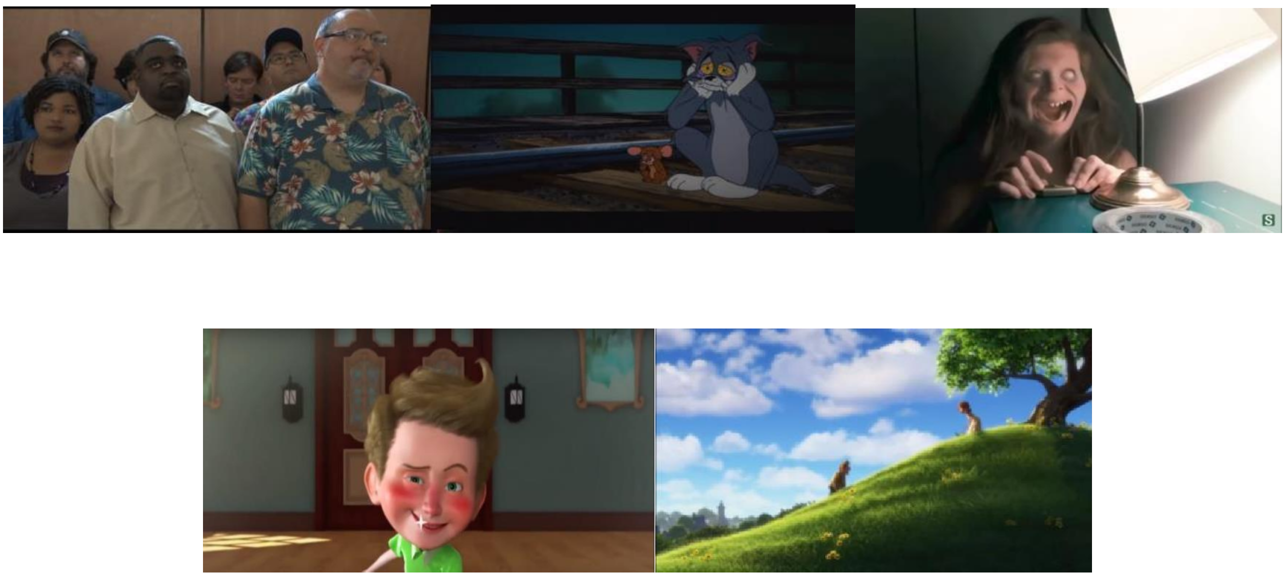


Fig.19: Screen shots of Videos Selected for Experiment

These videos were used in order to observe and analyze the sequence and variations in every participant's emotion.

4.5.2 Interviewing Environment

In order to achieve efficient stimulation, it was decided to book a meeting room where participants could be interviewed. This allowed us to add up another question whether surrounding physical environment play role in stimulation or not? For this task two different interviewing environments were arranged. The first interviewing environment was set up in a projector room with speakers which assisted in delivering efficient stimulation to the participants. The projector room was made dark, the screen brightness and speakers volume were adjusted which suited normal human condition. About 25 participants were interviewed in this enclosed environment. The second environment involved more manual work; the videos were shown through mobile device with earphones. Fig.20 illustrates the environment set up for same. In this case participants were interviewed in an open environment. The system was successful in detecting the facial expression and was able to predict the participant’s emotions. As long as the background colour was normal and the system was able to pin point the methods from the ‘detection’ part, the system performed without any hesitation.

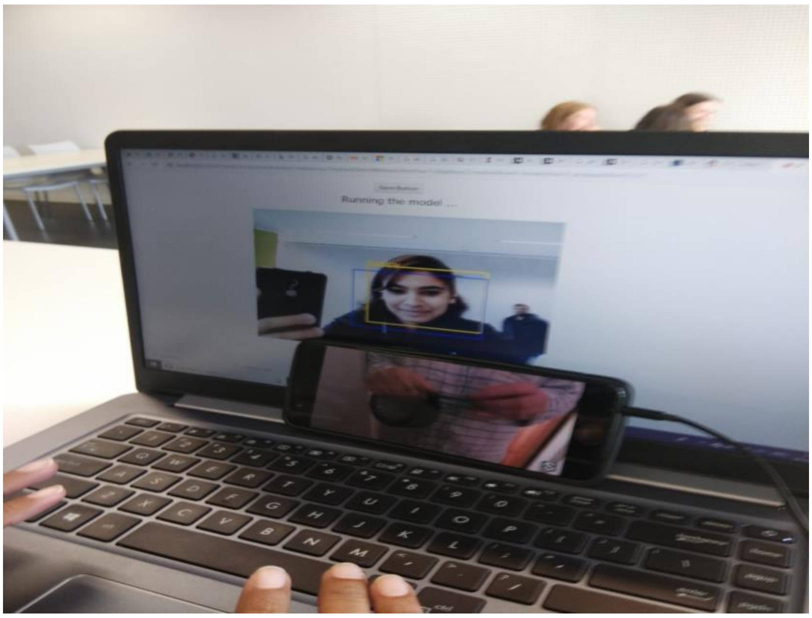


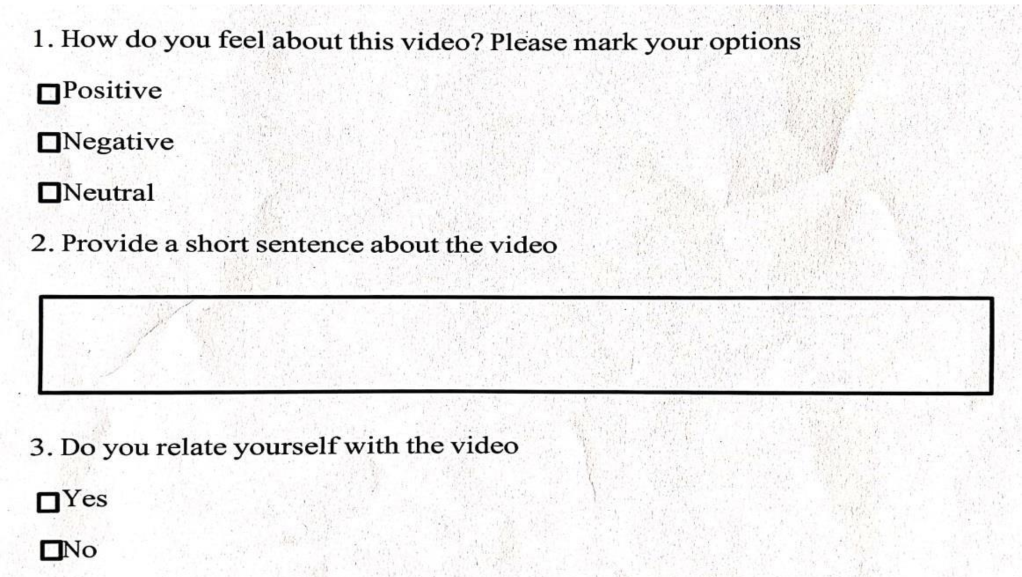
Fig.20: Environment Set-up for Interview

The results obtained from both the interviewing environment do show slight variation. In case of horror videos, high emotional count was observed from the second interviewing environment than the first one. This may be due to the fact that earpiece were used in delivering the stimulation. This clarifies one thing that surroundings to play important role if the stimulation is delivered in the right way and in right amount.

5. Data Acquisition

In this experiment there were two forms of data's which were collected from the participants. The first dataset involved utilization of computerized systems for stimulation and capturing of emotions. It holds relative information of the participants like first and last name, gender, nationality and age, followed by emotions which were captured during the course of the interview. The data of captured emotions along with relative information are stored in a text format as illustrated in Fig.21**.** This dataset allowed in identifying the number of times a particular emotion occurred for a participant. Later, emotions from this dataset were converted to sentiment in order to identify the correlativity between the system and the participant's self-evaluation to check how accurate the system was.

After the completion of a video, participants were asked to manually fill-in a self-evaluation form, in which they were asked to fill-in their emotional perception and feelings with respect to the video viewed to them. In short, participants were asked about their emotional feeling which they felt when seeing the video. This formed the second dataset in which three questions were framed as in the Fig. illustrated below. Every participant was well informed about what kind of information was required from their end on this from.


1. How do you feel about this video? Please mark your options

☐Positive

☐Negative

☐Neutral

2. Provide a short sentence about the video

3. Do you relate yourself with the video

☐Yes

☐No

Fig.21: Self-Evaluation Form

The questions from the second dataset were taken in as attributes (Q1, Q2, and Q3) for further relational and geospatial analysis to clarify our hypothetical questions.

6. Data Analysis

The baselines for emotions captured by the system were different for every individual participant. These baselines were on the criterion of facial expressions detected by the system. The types of emotions captured and their number of occurrence were noted for every participant, this allowed in identifying which particular emotion was experienced the most by an individual during the course of interview. Also, certain types of emotions were recorded in large numbers by the system for some participants, this could be easily observed from the illustration in **Fig.22** which shows some participants had high emotional count. From this we can make an observation that majority of the participants were emotionally strong, but this does not hold the fact that other participants are emotionally weak.

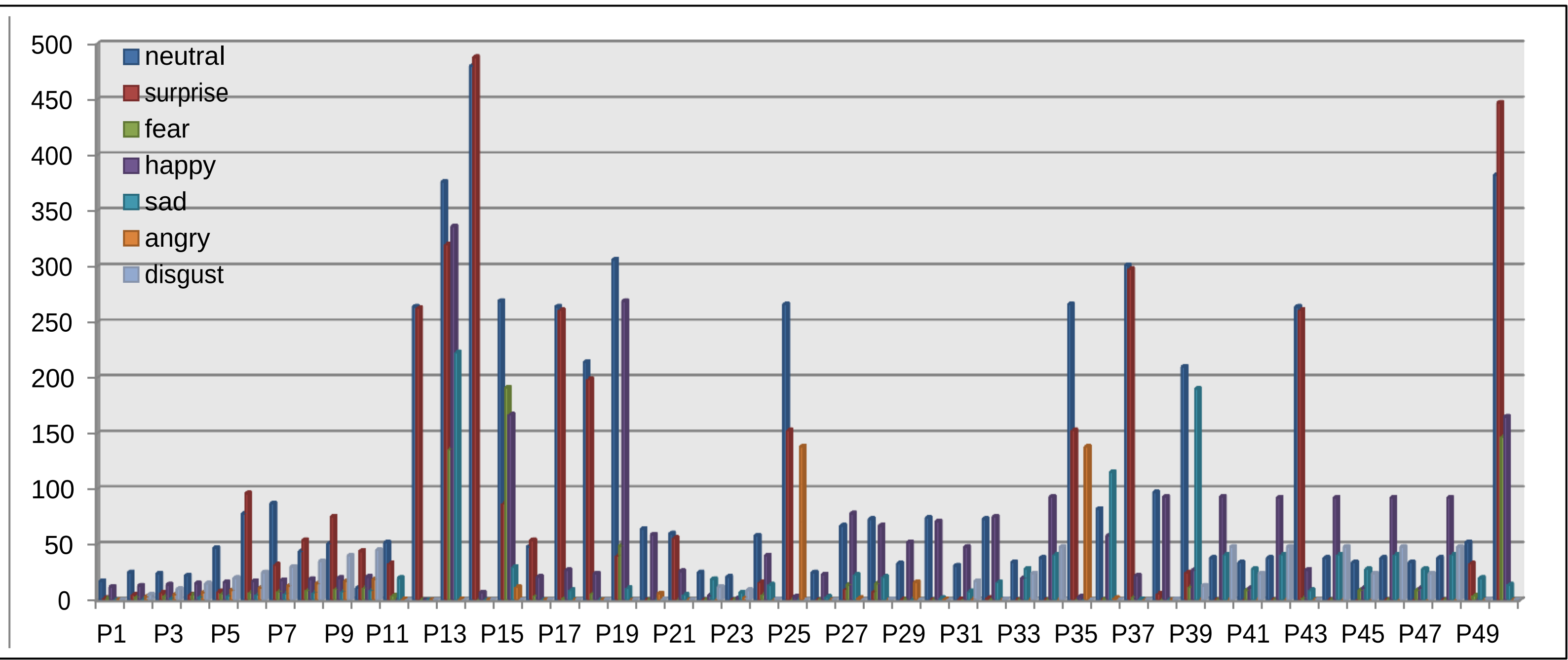


Fig.22: Overall Emotional Count For 49 participants

The emotions were also analyzed on the basis of video types, i.e. from a particular video type which participant's had highest emotional count. Further, the emotional count obtained from the system which formed the first dataset were converted into sentiments in which emotional values were assigned as, neutral = (neutral); positive = (happy, surprise); negative = (fear, sad, angry and disgust). This was done in order to calculate the correlativity between the system and participants (self-evaluation form). This conversion of emotions into sentiments produced decisive results in which the positive emotion count were comparatively more than negative and neutral.

The second dataset (self-evaluation form) were manually processed into digital format in R-studios, which allowed processing, filtering and visualizing of data in many possible ways.

Here, sentimental responses with respect to videos; gender based sentimental feedback; response to videos with respect to attribute Q3 and participants by nationality were analyzed. This dataset also assisted in identifying basic info like, number of both male and female participants, how many viewed which videos and number of positive, negative and neutral feedbacks from the participants.

Filter

	Nationality	Gender	Age	Video	Q1	Q2	Q3
1	German	Male	23	Elevator(Funny)	Positive	Funny at the start with quite an interesting twist	Yes
2	German	Male	22	Elevator(Funny)	Neutral	Don't take the elevator	No
3	Nepali	Male	28	Elevator(Funny)	Negative	Change can't be good everytime	Yes
4	Nepali	Male	25	Elevator(Funny)	Positive	Being clever is not always good	No
5	Nepali	Male	24	Tom&Jerry(Sad)	Neutral	it was quite boring	No
6	Indian	Fe-mlae	27	Horror	Negative	Unacceptable ending, it was very scary	Yes
7	Indian	Male	24	Plank(Funny)	Positive	It was funny, happy moments occur in life	No
8	Cameroon	Male	27	Elevator(Funny)	Neutral	Shows two sides of the story	No
9	Indian	Male	25	Elevator(Funny)	Neutral	Funny but also felt sad for the subject might suffer with...	No
10	German	Male	22	Tom&Jerry(Sad)	Negative	sad, hearbreaks and lost hopes	No
11	German	Male	23	Horror	Positive	I kinda like these kind of videos. They are funny but nor...	Yes
12	Indian	Male	21	Tom&Jerry(Sad)	Neutral	Seen the video before, it is always amusing	No
13	German	Fe-mlae	20	Elevator(Funny)	Positive	I felt empathic with the small guy in the elevator but it w...	No
14	German	Fe-mlae	19	Tom&Jerry(Sad)	Neutral	Video about Tom & Jerry, but this time a rather sad story...	No
15	German	Fe-mlae	20	Plank(Funny)	Positive	his facial expression was a bit disgusting can't explain it ...	No
16	Pakistani	Male	27	Plank(Funny)	Positive	Pretty humorous, when the kid was blushing due to fru...	No
17	German	Male	21	Horror	Neutral	The video was ok,a little bit scary but overall relatively n...	No
18	German	Fe-mlae	21	Plank(Funny)	Positive	It was funny and easy to understand and comedic timin...	Yes
19	Italian	Male	22	Tom&Jerry(Sad)	Negative	Sad but relatible story with funny jokes included	Yes
20	Dominican Republic	Male	26	Plank(Funny)	Positive	It was funny	No
21	vietnam	Fe-mlae	28	Horror	Negative	Just scare in the final scene	Yes
22	Indian	Male	26	Tom&Jerry(Sad)	Neutral	A little amusing at the end. Amused at the length to whi...	No
23	India	Male	26	Tom&Jerry(Sad)	Neutral	The video was not that emotional at all	No

Showing 1 to 23 of 49 entries

Table.1: Second Dataset Imported in R-Studios

Later, the attribute Q2 from the second dataset were analyzed with the lexical approach, to identify which lexicon is efficient in determining the emotional valiance. Text-mining approach was followed on the attribute Q2 in which participants were asked to express their emotional feelings in from of sentences. By using various lexical approaches like NRC, BING, AFINN, tidytext and Syuzhet, emotional as well as sentimental tone of the participants were analyzed.

The amount of data collected from both the sets contained large amount of information which was processed and viewed in multiple ways as possible. The observations made out of these datasets helped in answering the hypothetical questions and understanding the futuristic outlook for this project.

7. Results

This section is categorized into three sections which basically answers the derived hypothetical questions as stated in **Section 1.2**. Some other interesting results were also obtained which are also discussed in this report. The results attained through these investigations are completely objective and has not been compared with any other related works.

7.1 Correlation

Correlation in general is defined as a measure of the linear relationship between two quantitative variables. In this experiment, the correlation of sentiments between the system and self-evaluation form filled in by the participants has been determined. Now in correlativity, when a value of one variable increases as the values of the other increase, it is known as positive correlation. On the other hand when value of one variable decreases as increase in value of another variable; it forms an inverse relationship and is known as negative correlation.

There are many correlation techniques, but the one which best suited this experiment was the '*Pearson's correlation coefficient*', which measures the strength of two variables in relation to each other. The Pearson correlation formula produces an outcome which is referred as coefficient value which ranges in between -1.00 to +1.00. If the coefficient ends in negative range it means that the variables are negatively related to each other. Similarly, if the coefficient ends in positive range, the variables are positively related to each other.

The Pearson Coefficient is actually named as the Pearson Product Moment Correlation (PPMC), the formula is discussed as below:

$$Correl\,(X,Y) = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where:

n = number of scores

Σxy = sum of product of paired scores

Σx = sum of x scores

Σy = sum of y scores

Σx^2 = sum of squared x scores

Σy^2 = sum of squared y scores

Fig.23: Pearson Correlation Coefficient

In our experiment, we used the *Correl* function in excel to determine the final correlativity. [8]

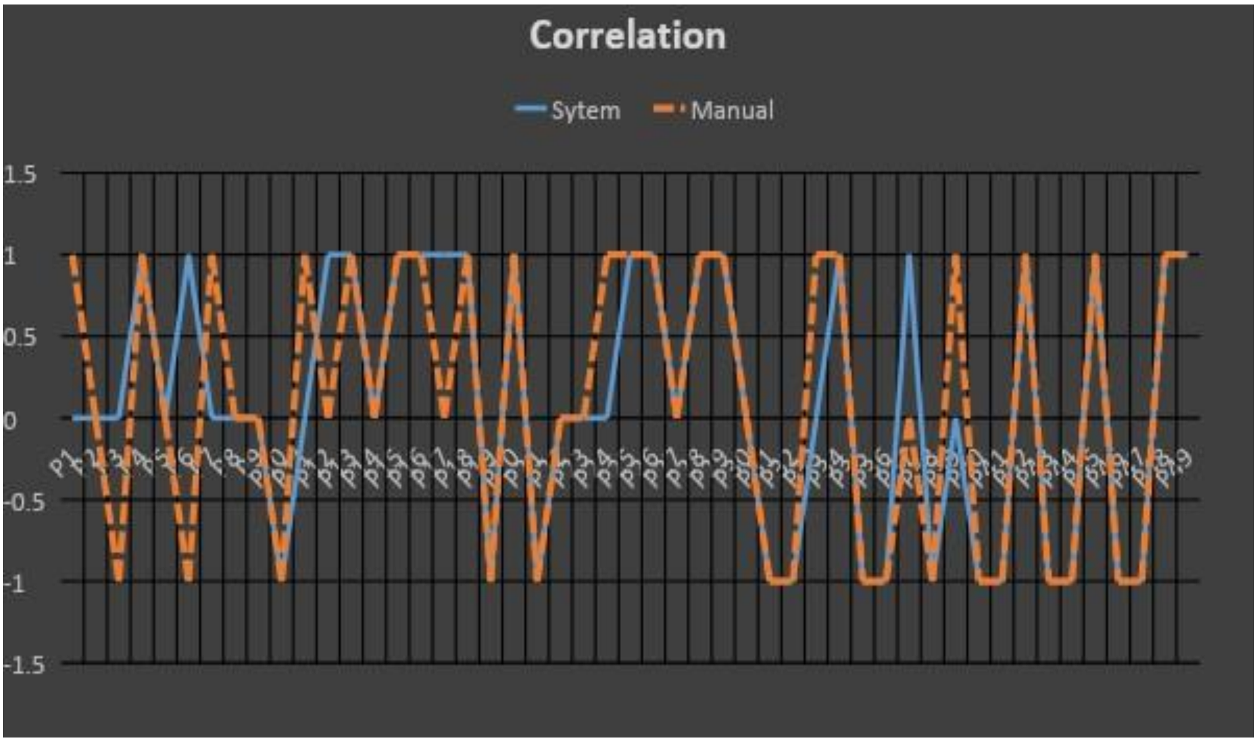


Fig.24: Correlativity of sentiments between the System and Self-evaluation (Manual)

The above illustrated correlation in line graph format, we see that the colour represented as blue is for system recorded data, orange dotted line for manual recorded data and the combination is colored is the similarity in both data. In above analysis the Y-axis is divided mainly in three scale as 1,0,-1 to represent positive, neutral and negative sentiments of the candidate. The graph depicts the result of 49 candidates with a correlation of **0.798384886**, which is **79%** and is on the positive range. With other observation noted from the obtained results (discussed in later sections) this shows that the system is working correctly and can be used as a system for sentimental as well as emotional analysis.

7.2 Cross-Cultural Background

The observations and results obtained in this section provided support to the proposed theory that emotions and sentiments can be similar among humans irrespective of their ecological and cultural background. Even though hailing from different parts of the world and speaking different languages, the sentimental or emotional aspects of humans are similar to each other to certain extent. About 49 participants from the university were interviewed hailing from different countries. The **Fig.:** below give a breakdown of number of participants from different countries both male and female.

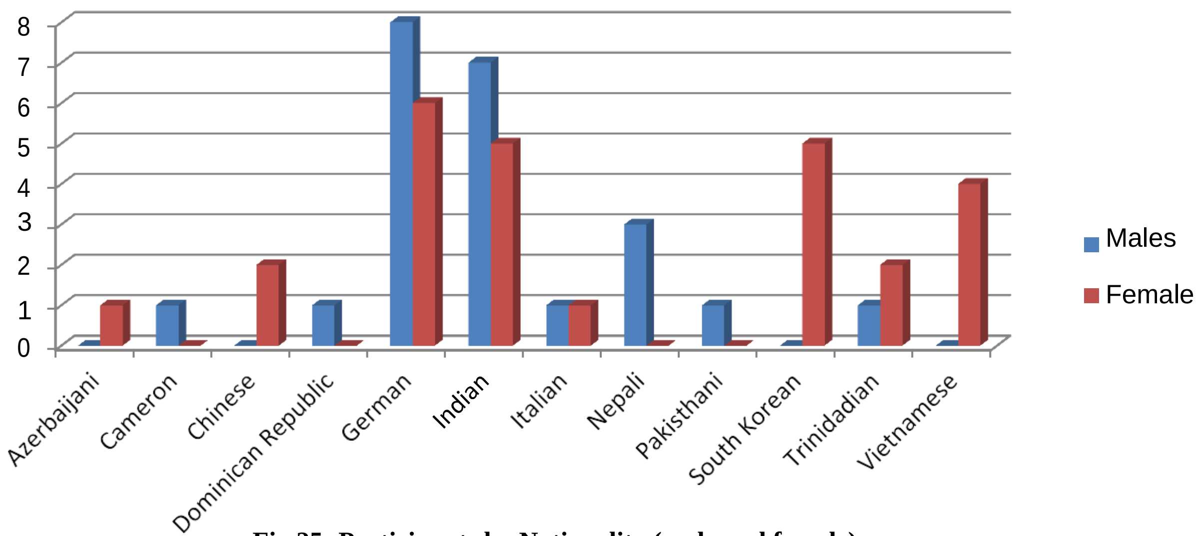


Fig.25: Participants by Nationality (male and female)

From the above presented graph, we can observe that majority of the participants were Germans (8-male and 6-female), followed by Indians (7-male and 5-female) and then South-Koreans (5-female). On the other hand only few participants were interviewed from Azerbaijan, Cameron, Dominican Republic and Italian, this was because participants were interviewed on basis of their availability.

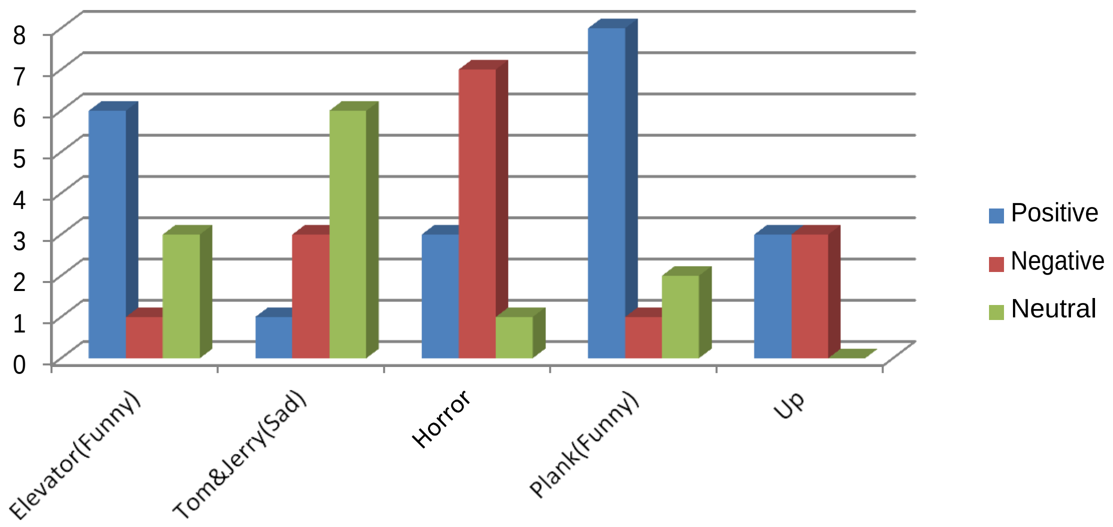


Fig.26: Sentimental Responses with Respect to Videos (male and fe-male)

The above illustrated graph represents the sentimental breakdown of all the participants with respect to videos. Note that these sentimental analyses are in line with the emotions detected by the system. Further, five different selected videos were randomly viewed to participants and their sentimental reading were noted. We can observe from first video, about six people gave positive response from which two were Germans and one each from India, Nepal, South-Korea and Trinidad, while one negative response from a participant hailing from Nepal. When the emotions of this particular participant were being analyzed, certain set of negative sentiments were observed which gives a proof that system recorded the emotions accurately and in-line with the participants response in self-evaluation form. Further, on the same video neutral responses were observed from each participant hailing from India, Cameron and Germany. In similar way responses of participants were observed for other videos.

This conclusion formed the first part of proof that irrespective of ecological, cultural and linguistic background, emotions/sentiments is something which is commonly perceived by most humans in common. To solidify this proof, data was analyzed with a different perception which is illustrated in Fig.27.

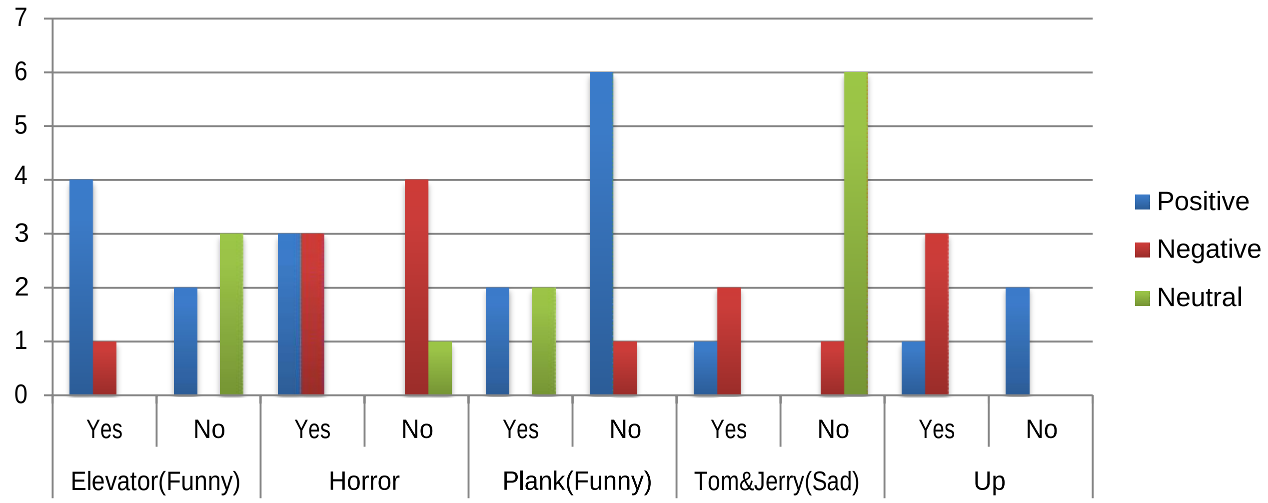


Fig.27: Sentimental Response with respect to attribute Q3

In this graphical presentation, data was analyzed with respect to attributes Q1 and Q3 taken from the self-evaluation form. The attribute Q1 consists of the sentimental responses of the participants where as Q3 states: Do you relate yourself with the video? In which participants were asked if they relate themselves with the situation presented to them in the interview, like has it ever happened to them in their life. Binary options in form of Yes and No were provided and the participant had to select one of either of these options. The responses obtained on this part of the investigation were quite satisfying.

From Fig.27 when considering the second video (Horror), there are in total eleven participants, out of which six participants relate themselves with the scenario which had happened to them in reality, but their sentimental responses are different (3-positive and 3-negative), while five participants do not relate themselves with the scenario and sentimental distribution is as (4-negative and 1-neutral). In a similar way the distribution of participant responses for the presented scenario are laid out in the graph (Fig.27). One can also observe from the graph that some participants had similar sentimental perception. If considering the response '**No**' for the third and fourth videos, in each, six participants have the similar feelings with respect to the scenario presented. This formed the second proof for cross cultural analysis and confirms the proposed theory that emotions and sentiments can be similar among humans irrespective of their cultural, ecological and different linguistic background.

The graph in Fig.27 also assisted in answering our third hypothetical question, in which participants do emotionally connect and relate themselves with the scenario in the video. From the graph one can easily notice the number of participants sentimental response with respect the attribute Q3 in terms of Yes and No. The following table can give a clear idea about the participant's response to attribute Q3 with respect to sentiments.

Videos	Q3	Positive	Negative	Neutral
Elevator(Funny)	Yes	4	1	0
	No	2	0	3
Horror	Yes	3	3	0
	No	0	4	1
Plank(Funny)	Yes	2	0	2
	No	6	1	0
Tom&Jerry(Sad)	Yes	1	2	0
	No	0	1	6
Up	Yes	1	3	0
	No	2	0	0

Table.2: Participant response to Q3 w.r.t sentiments

7.3 Lexical Approach

During the initial phase of the experiment, it was observed that an extensive amount of research has been done on sentimental analysis using the text mining approach. From the referred related works it was decided to identify which particular lexicon would be an efficient if the system is fully automated, that is, the entire manual analytical task could be done by the system itself. Though this task may be for long run but it did helped in making certain observations. From the self-evaluation form, the attribute Q2 which consisted of participants emotional feelings in sentence format were used in finding the answers to the fourth hypothetical question. The attribute Q2 was imported in R-studios where various lexicon libraries like **NRC**, **BING**, **AFINN**, under **tidytext** and **Syuzhet** packages were used along with their functions to analyze data with text-mining and lexical approach.

7.3.1 Finding Sentiments by Words

The first task of using the lexical approach was to determine the number of both positive and negative words from the attribute Q2 of the self-evaluation form. In lexical vocabulary the lists of words are distributed in positive and negative format, hence the neutral part has not been considered in this section of the investigation. The dataset were imported into R-studios in which unwanted variables were removed leaving only the attribute Q2 as only variable. Every sentence from the attribute Q2 formed one row each. By using functions like **mutate()** and **anti_join()** along with list of **stop_words** from **tidytext** package, the sentences were broken down into words, where every word from the attribute Q2 formed a row. [9] Using this approach, the group was successful in identifying the most commonly used positive and negative words by the participants in the attribute Q2 of the self-evaluation form. The Fig.28 illustrates the result accordingly.

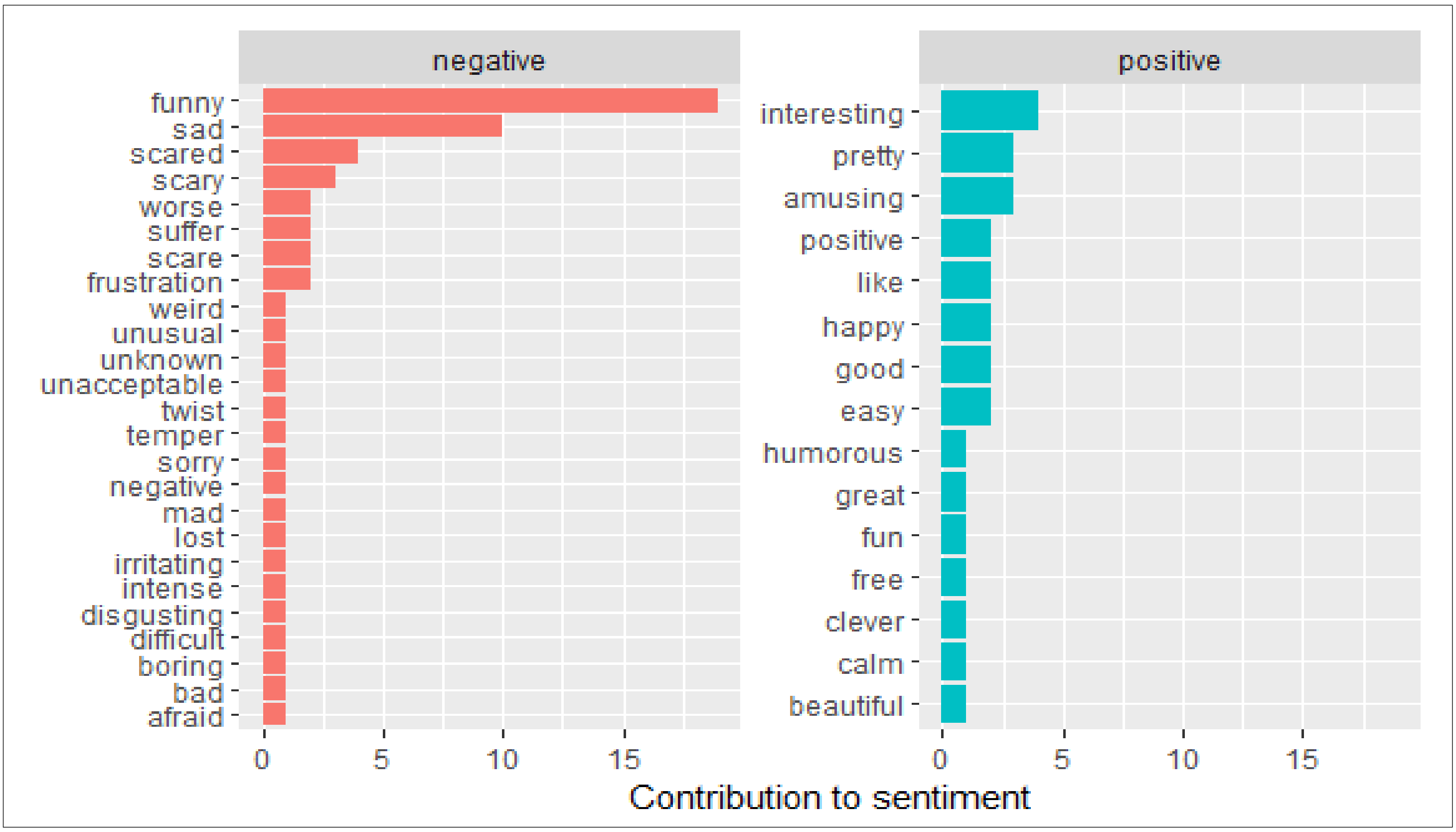


Fig.28: Commonly used Positive and Negative Word

The advantage of having such kind of frames with both types of sentiments and words is to analyse the word count contributing to the sentiments. With the help of both these arguments we can identify how much each word has contributed to a particular sentiment. From the above illustration, we can make out a peculiar observation, the word ‘Funny’ is categorised into negative sentiment, while the word ‘fun’ is in positive sentiment. This is because of the lexical vocabulary, as stated earlier that Lexicon are libraries which consists of list of words.

In this the **Bing** lexicon used did not contain the word ‘funny’ in its list of words hence it led to results as shown above. The other observation which can be made out, that more number of negative words were identified than positive words. Note that these results are obtained on basis of words contributing to sentiments for all 49 participants. This motivated in moving towards identifying sentences contributing to sentiments.

7.3.2 Finding Sentiments by Sentences

Syuzhet Package along with **get_sentiment()** function was utilized to achieve results for sentiments on basis of sentences. The attribute Q2 was used as the vector of sentences and all four methods (syuzhet, Bing, afinn and NRC) as the argument. Different methods utilized here reflected distinctive results because every method has a different scaling. Once the sentiment values were determined by each method they were summed up individually to get an overall emotional valence from the text (attribute Q2). The results were; syuzhet= 8.25; Bing= -32; afinn= 53 and NRC= 2, which clearly shows that the emotional valence were more towards the positive side. Further, the results were summarized using the **summary()** function, which provided an extensive distribution of sentiments from the sentences.

```
> summary(syuzhet_vector)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -1.750  -0.500   0.200   0.165   0.750   2.850
> summary(nrc_vector)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -4.00    0.00    0.00    0.04    1.00    2.00
> summary(bing_vector)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -4.00   -1.00   -0.50   -0.64    0.00    2.00
> summary(afinn_vector)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -5.00   -2.00    0.50    1.06    4.00   10.00
>
```

Fig.29: Summary statistics of

The above figure illustrates the distribution of sentiments for every sentence for every method used. Such comprehensive measurement of sentiments can be very informative, but they do not tell us the flow of sentiments for every sentence i.e. how the positive and negative sentiments are active across the sentences (attribute Q2). We can illustrate these measures by plotting values in a graphical structure where X-axis represents the number of participants and Y-axis measures the intensity of both positive and negative sentiments. The graphical structure for each method has been illustrated separately since the scaling for every method is different.

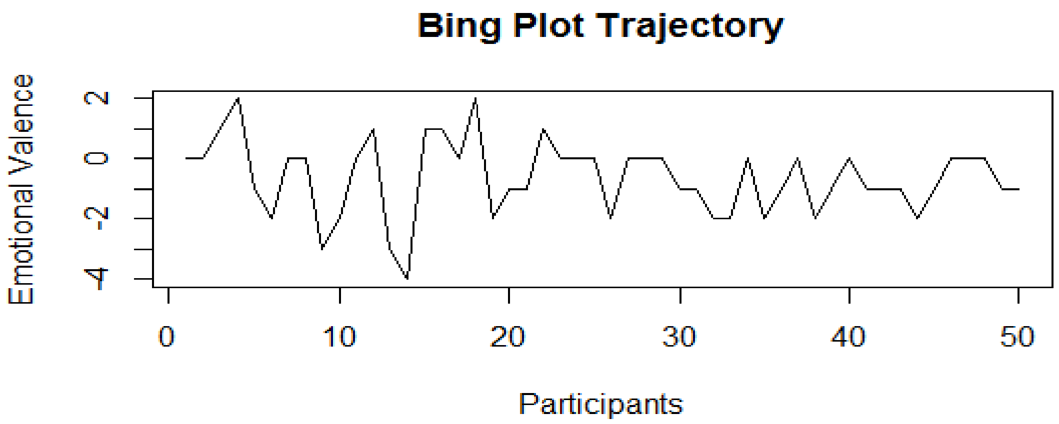


Fig.30: Sentiment Valence Using Bing Method

The trajectory starts from the neutral territory, moves to positive and then enters a period of negative to neutral and then again negative sequence. At sentence 14, the trajectory bends towards the negative side (-4), and recovers back to 1 in the later sentence. Majority of the trajectories are towards negative side. Mean value = -0.64

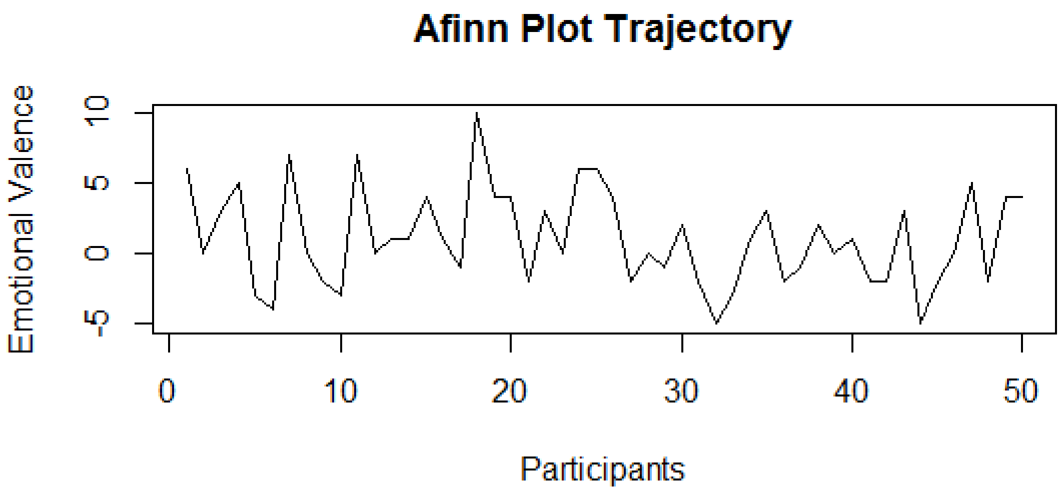


Fig.31: Sentiment Valence Using Afinn Method

Here, the trajectory starts from the positive side and moves to neutral state. At sentence 18, the trajectory reaches the peak of the graph and the lowest valance is recorded at sentence 32. Majority of the projections are towards the positive side with Mean Value = 1.06.

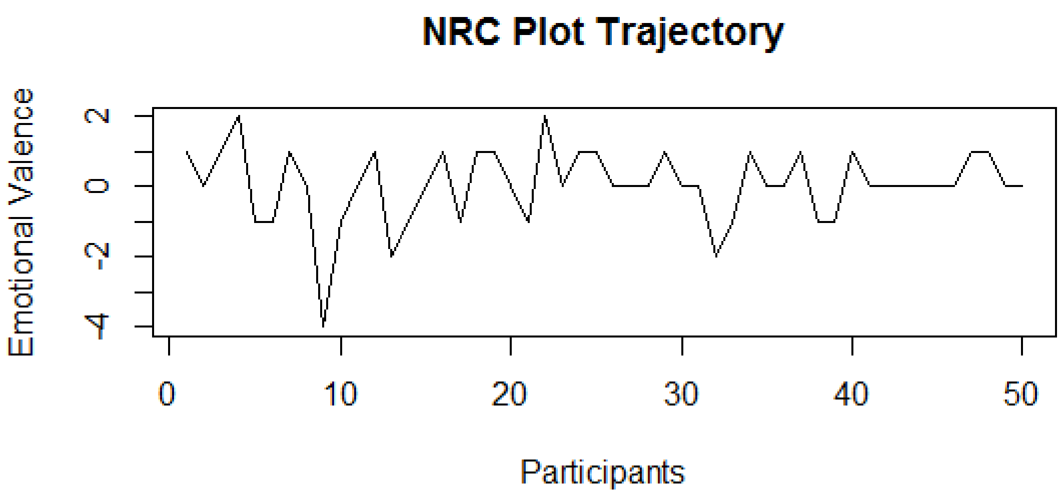


Fig.32: Sentiment Valence Using NRC Method

Here, too the trajectory starts from the positive side and moves to neutral phase. At sentence no. 9 the projection moves to the lowest phase at -4. On close observation, it can be noted that majority of the trajectory are moving from neutral to positive phase. From sentence 22 to 31 the projections are between 0 to 1 and similar projectiles can be observed from sentence 40 to 49. Mean value = 0.04.

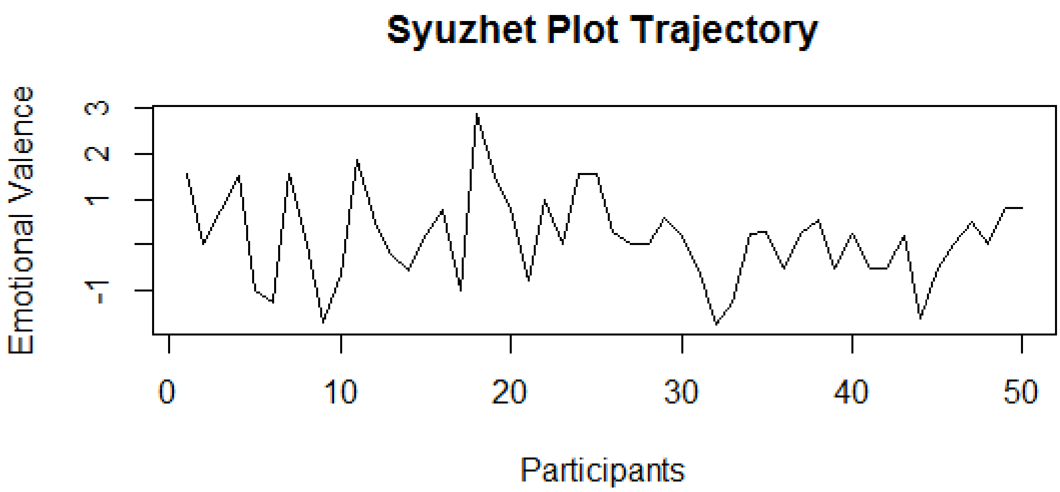


Fig.33: Sentiment Valence Using Syuzhet Method

In Syuzhet, the trajectories are mostly between neutral and positive section. The minimum value can be observed at -1.75 at sentence 32. The maximum value for the projection is at sentence 18 with 2.85 as the valence value. Mean value = 0.2.

On observing these trajectories, one can notice the difference in the scaling's of each method. Also the sentiment values of the attribute Q2 for every method is different. These observations are useful for demonstrating that each lexical approach is different and that there is no such standardization which defines which particular lexicon can or should be used in sentimental and emotional analysis on global scale. Further, the emotional and sentimental values for every sentence was determined by using **get_nrc_sentiment** which implements **NRC** lexicon in which every row is converted into sentence from the source file, each column consists of every emotion type along with sentiment valence: positive and negative. The emotional and their values along with sentiment values are illustrated below

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
17	1	2	1	2	0	1	1	0	1	0
18	0	0	0	0	0	0	0	0	0	1
19	0	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0
21	1	1	0	1	0	0	1	0	1	0
22	0	0	0	0	2	0	0	0	0	2
23	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	1
25	0	0	0	0	0	0	0	0	0	1
26	0	0	0	0	0	0	0	0	0	0
27	1	1	1	0	1	0	0	1	1	1
28	0	0	0	0	0	0	0	0	0	0
29	1	0	1	1	1	2	0	0	1	2
30	0	1	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0
32	2	0	1	1	0	1	0	0	2	0
33	0	0	0	1	0	0	0	0	1	0
34	0	1	0	0	1	0	0	0	0	1
35	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0
37	0	2	0	0	1	0	0	1	0	1
38	0	0	1	0	0	0	0	0	1	0

Table.3: Emotion and Sentiment Values of Sentences

In order to calculate positive and negative valance for every sentence, the values from negative columns are converted to negative numbers and further added to positive columns; **valence <- (nrc_data[, 9]*-1) + nrc_data[, 10].** Further, these values are plotted graphically to give a visual count of both emotional and sentiment values separately.

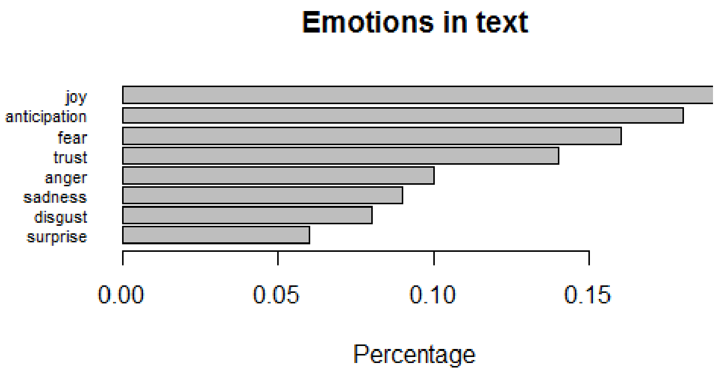


Fig.34: Emotion Count (get_nrc_sentiment)

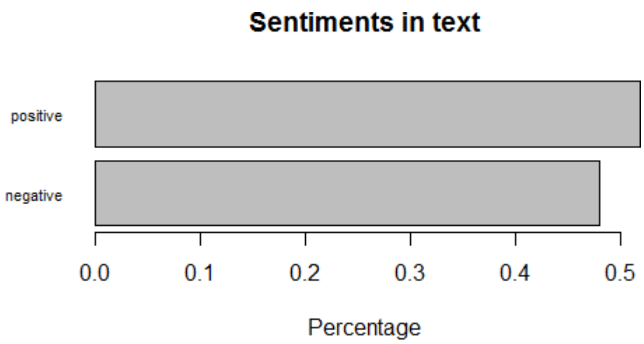


Fig.35: Sentiment Count (get_nrc_sentiment)

From the above plots, it can be observed that the percentage valance of positive emotions is higher than the negative emotions. The emotion and sentiment values can be visualized in various different means by using the **NRC** lexicon.[10] With reference to literature works referred much research is still needed on the lexical approach, as they are affected by factors like cross cultural domain and introduction of new words in the vocabulary. Further, there are no lexicon which can identify emotions and sentiments from other languages except English. The results obtained from the above analyses, one can opt for **NRC**- lexicon under **Syuzhet** package as they allow visualizing the emotion and sentiment data in different ways, followed by **Afinn** under **tidytext** as it is widely used by organizations in opinion mining over various platforms.

8. Discussion

The first stage of this project was to design and implement system which could capture emotions of participants. By integrating various datasets (**FER+** & **RAF-DB**) along with facial and emotion detection api's (**face-api.js** over **tenserflow.js**), the system was successfully implemented over a web-browser. Implementing such system allowed it's accessibility on mobile devices as well.

Few constraints were noticed during the system testing phase:

1. The model when run's occupies large amount of space in RAM which causes the model to crash. The model was later moved on to laptop which had higher configuration (16GB RAM) and it ran without any hindrance. Though the model can run on systems with lower configuration (8GB RAM) too, but will not deliver performance which one would expect.
2. Objects covering the face such as spectacles or cap will not allow predicting accurate emotions of participants.
3. Distance between the participant and the system should be nearly 30cm - 35cm, to capture the emotions accurately.
4. The face should be right in-front of the system; movement of head will not allow the system to capture the emotions accurately.

The self-evaluation was paper based; this can be converted into a web form just like the registration page over web-browser. This will reduce the manual work of converting the data into digital format, plus the system can directly analyse those feedback too.

The observations made from the above results do indicate that there is still scope for improving the human interaction with the system. The correlativity between the system and self-evaluation (manual) came out to be 79%, which is a good result. Presently, existence of such kind of system was not known with which accuracy of the systems could be compared. Further, the observation from results shows that emotions and sentiments are common among humans to a certain level, irrespective of cultural and ecological background; this could be on the basis of level of perception of participants. Some participants even related themselves with the situation presented to them during the interview; this shows that participants who relate themselves with the presented situation have similar emotional and sentimental responses.

In last, lexical approach was used to analyse the attribute Q2 from the self-evaluation form. Various lexicons (Bing, Afinn, and NRC) were used under Syuzhet and tidytext packages which allowed visualizing of the datasets. These lexicons consists of list of words from English vocabulary, hence a great amount of research is still required on the lexical approach in order to have lexicons from different languages vocabulary. As per the lexical approaches used in this project, NRC lexicon made by Saif Mohammad [11] along with Syuzhet package is recommendable as it allows analysing and visualizing both emotions and sentiments in different ways.

9. Conclusion

The desired objectives from this project have successfully been achieved. Applying the audio-visual approach, emotions were successfully derived from the implemented modelled system. Through self-evaluation correlativity and accuracy of the system were determined. Though there is scope for improving the interaction of the system with humans, the observation made by conducting this experiment has answered all the framed hypothetical questions. By making such systems more sophisticated, it can be utilized by various organizations to understand the sentimental and emotional valence of people or employees and help in identifying and diagnosing psychological disorders including fating, stress, anxiety and depression at their very early stages.

10. References

Kendra Cherry (2019) Overview of the 6 Major Theories of Emotion,Available from: <https://www.verywellmind.com/theories-of-emotion-2795717>, [Accessed 23 June 19].

Nada Allouch (2018) Sentiment and Emotional Analysis,Available from: <http://blog.emojics.com/emotional-analysis-vs-sentiment-analysis/>, [Accessed 22 June 19].

Taboada, Brooke, Tofiloski,Voll & Stede (2010) Lexicon-Based Methods for Sentiment Analysis,Semantic Orientation CALculator, Volume 37 Number 2,Canada: Simon Fraser University.

Kucher et al. (2018) The State of the Art in Sentiment Visualization: Sentiment Visualization Techniques, Volume 37, Number 1,Sweden: Linnaeus University.

Ohman, Honkela & Tiedemann (2016) The Challenges of Multi-dimensional Sentiment Analysis Across Languages, Finland: University of Helsinki, pp.1,3,4.

Sauter et al.(2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations,Volume 107, Number 6, pp.2,3,5.

MindField BioSystem (2018) Introduction to Audio-Visual Stimulation, Available from: <https://www.mindfield.de/en/AVS/Introduction-to-Audio-Visual-Stimulation-AVS.html>, [Accessed 23 June 19]

Stephanie Glen (2019) Statistics How To: Correlation-Coefficient, Available from: <https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/correlation-coefficient-formula/>, [Accessed 25 July 19]

Julia Silge, David Robinson (2019) Text Mining with R, 1st ed., Place of Publication:Massachusetts,United States ,Publisher:O'Reilly Media.

Matthew Jockers (2017) Introduction to the Syuzhet Packaage, Available from: <http://blog.emojics.com/emotional-analysis-vs-sentiment-analysis/>, [Accessed 7 May 19].

Saif Mohammad (2017) NRC Word-Emotion Association Lexicon, Available from: <http://saifmohammad.com/WebPages/lexicons.html>, [Accessed 17 June 19].

Bing Liu & Minqing Hu (2004) Opinion Mining, Sentiment Analysis, and Opinion Spam Detection, Available from: <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>, [Accessed 16 June 19].

HOYEOL KIM (2018) Limits of the Bing, AFINN, and NRC Lexicons with the Tidytext Package in R, Available from: <https://hoyeolkim.wordpress.com/2018/02/25/the-limits-of-the-bing-afinn-and-nrc-lexicons-with-the-tidytext-package-in-r/>, [Accessed 16 June 19].