Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Investigating the significance of head orientation in Automatic Emotion Recognition Systems

Yatheendra Pravan Kidambi Murali

A Dissertation

Presented to the University of Dublin, Trinity College

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science (Intelligent Systems)

Supervisor: Dr. Khurshid Ahmad

August 2022

Declaration

I, Yatheendra Pravan Kidambi Murali, declare that the following dissertation, except where otherwise stated, is entirely my own work; that it has not previously been submitted as an exercise for a degree, either in Trinity College Dublin or in any other University; and that the library may lend or copy it or any part thereof on request.

Yatheendra Pravan Kidambi Murali

August 19, 2022

Investigating the significance of head orientation in Automatic Emotion Recognition Systems

Yatheendra Pravan Kidambi Murali, Master of Science in Computer Science

University of Dublin, Trinity College, 2022

Supervisor: Dr. Khurshid Ahmad

Non-verbal communication has multiple modalities. Gestures and head orientation encompass significant emotional cues. A case study analysis of 221 spontaneous videos is presented which evaluates the agreement of emotion recognition systems based on profession and investigates the relationship between head orientation and perception of emotions. These videos are of charismatic people (like CEOs and Politicians) or by people who represent them (Spokesperson). A semi-spontaneous video dataset is created, and the relationship of head orientation (represented through Euler Angles) with emotions identified by prominent emotion recognition systems(EMOTIENT FACET and AFFECTIVA AFFDEX) is investigated. The two systems have a good statistical agreement on the estimated head direction and the estimated Euler angles are highly correlated. There is a variation in the distribution of Euler angles between the systems and the distribution varies for different emotions. The relationship between head orientation and emotions is explored using regression analysis and fuzzification of Euler angles, and the results are in line with the evidence from the literature. Evidence for the association of specific head orientation for each emotion is found to be consistent within the system(intra-system) and varies between systems.

Acknowledgments

I take this opportunity to thank everyone who helped and supported me during this dissertation work.

Firstly, I would like to express my earnest gratitude to Professor Khurshid Ahmad, whose high energy levels, constant excitement, dedication, encouragement, knowledge, and the finest attention to detail have been an inspiration and were essential in making this thesis a reality. His in-depth knowledge and vast experience in the field of computer science and statistics helped me solidify my concepts and overcome the difficulties faced in this research.

Secondly, I am indebted to Professor Dr. Carl Vogel, for offering valuable advice on statistics and research methods. Thirdly, I would like to thank Deepayan Datta and Wanying Jiang for the collaboration during the data collection phase and brainstorming sessions.

A big thanks to Subishi Chemmarithil, who was kind enough to help us acquire data on certain politicians and spokespersons. A big credit goes to the Department of Computer Science and Statistics, Trinity College Dublin, for providing me with adequate infrastructure to work on my thesis. I would like to thank my parents K K Murali and V S Gayathri Devi, and my friend Bumija Sriraman, whose love and encouragement guided me throughout the entire MSc program.

Yatheendra Pravan Kidambi Murali

University of Dublin, Trinity College

August 2022

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Head Orientation and Emotion Recognition

Research advancements in the field of emotional intelligence and enhancement of capabilities of emotion recognition systems have increased by multiple folds in the last few years. These systems look at different facial landmarks to identify the face, discriminate the smallest changes in facial muscles to identify the facial expressions, and ultimately recognize emotions. However, other modalities like head orientation, variation in speech signals, and gestures carry important emotional cues. To improve the perception of emotions, these systems need to model all these non-verbal emotional cues and not look at the face in isolation.

There are multiple advancements to comprehend these non-verbal emotional cues. For instance, the automatic emotion recognition (AER) system considers the head as a rigid body, and the orientation of the head with respect to a reference axis is modeled using Euler angles. Figure 1.1 represents the three Euler angles that characterize the head orientation. The Euler angles provide a means to represent the head in a 3-dimensional space. The yaw angles capture the head orientation on the horizontal axis, pitch angles capture the head orientation on the vertical axis, while roll degrees present the information about the head tilt. These three Euler angles characterize the head orientation in the three-dimensional space. The head pose estimation from images and videos is an interesting problem in the domain of computer vision. Different systems use different approaches to solve this problem of head pose estimation where, the video is split into discrete frames and the head pose is estimated for each frame.

The primary objective of the presented study is to explore the impact of head orientation in automatic emotion recognition systems. A case study is presented where 221 spontaneous videos were analyzed to understand the relationship between head orientation and emotions. Videos of charismatic people addressing mass media are selected which include, speeches or interviews by CEOs of multinational organizations, Politicians, or people who represent them (Spokespersons). The curated dataset is semi-posed or semi-spontaneous because these individuals either have a script prepared or use a teleprompter, and also answer impromptu questions. Our data set comprises videos of individuals from 12 different nationalities, and the age range is between 28 to 91. The female to male ratio in the dataset is 97:124. Two different emotion

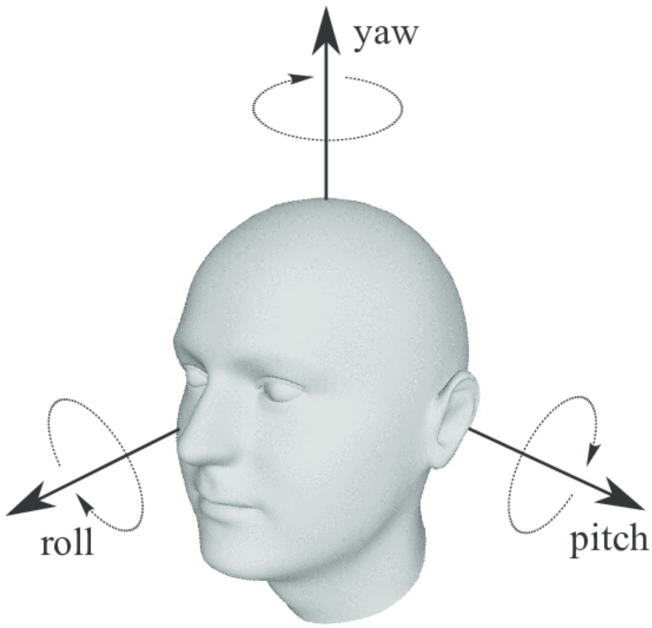

Figure 1.1: Illustration of Euler angles that characterize head orientation: Yaw, Pitch and Roll. Figure reproduced from (4)

recognition systems are used which detect the evidence for 6 basic emotions and estimate the head orientation for each frame synchronously. The correlation between the head orientation and the depicted emotion is examined with respect to these systems.

1.2 Research Contribution

The presented research is a part of a bigger research project which explores multiple modalities of non-verbal communication. The presented work, however, focuses on head orientation. Two popular emotion recognition systems (Emotient and AFFECTIVA) are considered which are capable of estimating the emotion evidence and head orientation for every frame synchronously. As part of this research project, 33 videos of CEOs are added to the collection of a massive dataset of semi-spontaneous facial expressions. The existing dataset is diversified in terms of nationality, gender, ethnicity, and the current research project adds a new dimension: Profession. The key contribution of this research is the development of a systematic methodology to understand the similarities and differences in these automatic emotion recognition systems with reference to head orientation. A detailed methodology pipeline is proposed to investigate the impact of head orientation on the perception of emotions by these emotion recognition systems. A extensive analysis is presented to understand the relationship between head orientation and the depicted emotions. Additionally, the presented analysis is extended by applying fuzzy logic which is still in an experimental phase and would provide a foundation for future work in the same domain. The significant results from the presented research is collated as a paper and submitted for the Future of Information and Communication Conference (FICC) 2023 (9).

1.3 Dissertation Structure

The rest of the dissertation is structured as follows: The motivation for the presented work with the review and summary of the related literature along with the research questions is discussed in chapter 2. The methodology pipeline designed to investigate the research questions along with the rationale behind selecting the specific statistical tests for hypothesis testing are detailed in chapter 3. Chapter 4 presents the hypotheses, experiments, and results in two folds. The initial part discusses the experiments conducted with reference to facial expressions (research question 1) and the later part discusses the experiments and presents the results with reference to head orientation analysis (research questions 2 and 3). The final chapter summarises the inferences made from the experiments and compares them with the evidences in the literature. Additionally, the future extension of the project work is discussed in detail. Additional tables and graphs are included in the appendix section for reference.

Chapter 2

Motivation and Literature Review

According to American Psychological Association (APA), nonverbal communication is defined as "the act of conveying information without the use of words". Non-verbal communication can complement, regulate, substitute for, or accent a verbal message (10). When a person emotes, he may provide non-verbal emotional cues either intentionally or unintentionally. (11; 12; 13; 14) indicate that the person may orient his head at a particular angle or adopt a specific gesture or modulate his speech signals to complement the communication of intended emotion. The related literature is explored to identify the potential gap in the literature and to support the motivation of the presented research.

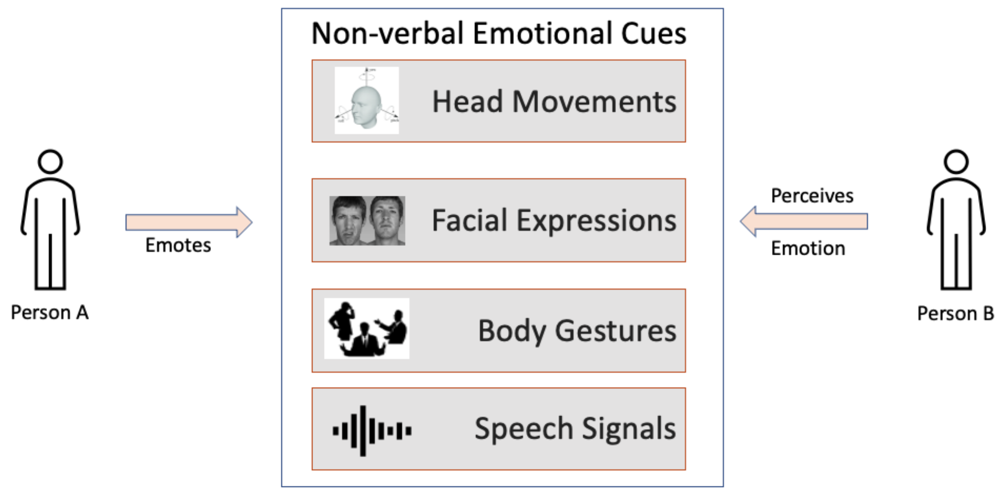

Figure 2.1: Non-verbal Emotional Cues

2.1 Facial Expressions

We can observe that, mostly when a person is angry they may lower their eyebrows or tighten their lips. Similarly, when a person is happy he/she may smile or raise his/her cheek. This presents a basic premise that there may be a set of discrete facial expressions associated with different emotions. People often exhibit these facial expressions to emote, either voluntarily or involuntarily (15). Charles Darwin was the first among the few to stipulate that a certain facial configuration is possibly the expression of certain emotion categories. (2). Darwin was a key contributor to the development of Facial expressions. Darwin looked at photographs and pictures from the works of French neurologist Duchenne De Boulogne and made keen observations on various facial expressions (16). Darwin made keen observations on different

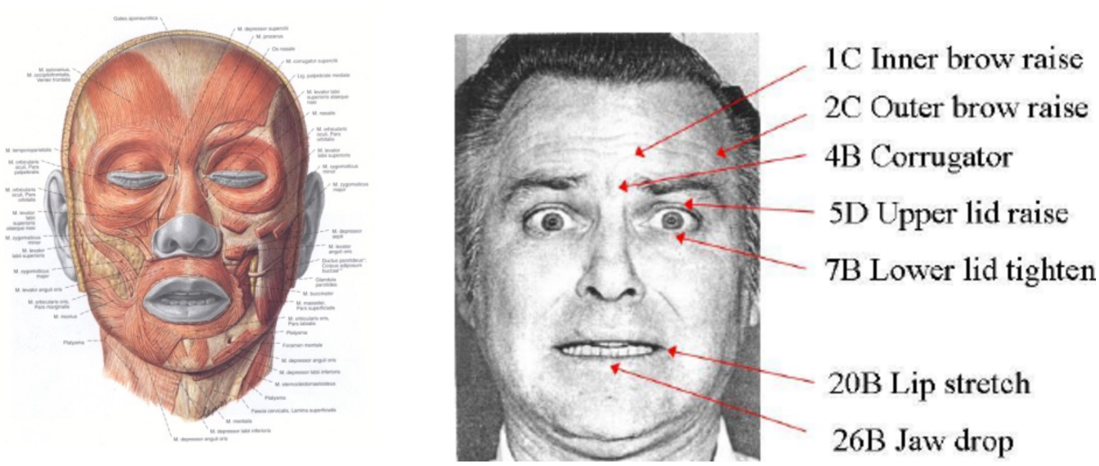

Figure 2.2: Anatomy of facial muscles and change in appearance during depiction of an emotion, figure reproduced from (1)

emotions and focused on changes in the appearance of facial muscles and the musculature of the face. The anatomy of facial muscles and the corresponding change in appearance during the episode of a particular emotion is shown in the figure 2.2. Darwin's major focus was on the face and he did not consider other modalities that enabled the perception of emotions. There are various contributions from the related literature (17; 18; 19) that indicates that there is a significant contribution from different modalities like speech, gestures, head orientation, head movements and facial expression may not alone convey the actual emotion (20). The other key aspect of the stipulations made by darwin is that, the timing of the expression was not considered, however, significant evidence highlighting the impact of facial movements (21) in the depiction and perception of emotions was presented in the literature later. Darwin's major observations were inspired by photographs and anatomy of facial muscles and thus, emotions were considered as discrete entities with no overlap whatsoever. Further, he also proposed that emotions are universal. Although many of darwins observations are widely accepted, most of the stipulations from darwin are heavily challenged in the subsequent literature. Paul Ekman, who is a pioneer in the field of facial expression and emotions, took inspiration from darwin's work and identified the gap in his stipulations. In (22), he presented arguments that emotions cannot be considered as discrete entities with strict boundaries and highlighted that the variations of the same emotions were not considered. Ekman used all these identified gaps to present the need for a systematic method for measuring facial movements and the inspiration behind developing the Facial action coding system. Ekman highlighted the need to understand the movement of muscles in terms of the smallest component unit called the action units and stipulated that the variation in a certain emotion can be understood with the difference in the interaction between such action units (22). FACS was initially developed in the year 1970 by Carl-Herman Hjortsjö. (23). Paul Ekman and Wallace Friesen used this as a basis and published a refined and revised version in the year 1978 with a goal to define any facial movement in terms of action units (1). They proposed a novel scoring methodology to systematically encode the facial movement and developed a formula based on inter annotator agreement. The authors, further developed this system in the year 2002. The modern systems are highly capable in terms of pattern recognition and FACS provides the basis for many modern automatic emotion recognition systems like Emotient (7) and Affectiva (24). Such systems integrate the foundation provided by FACS and utilize optimized pattern recognition algorithms to capture the smallest changes in the facial muscles to infer the implied emotion. FACS was influential in the development of basic emotion theory which theorized that, there are six discrete basic emotions. (25) proposed that there are six basic facial expressions of emotions, however, the concept of basic emotion theory is heavily challenged in literature and (26; 27; 28) characterized that there are four basic emotions in humans. On the contrary, (29) presented the idea that the exhaustive emotion set of 22 emotions can be created by compounding basic emotions. However, the idea of basic emotion theory was widely accepted and subsequent works have tried to explain the variation of emotions by considering them as a function of action units.

Emotion as a function of Action Units FACS provided a way to quantify the smallest facial movements in terms of action units. (30), explored the interaction of action units in different varieties of anger. The participants in the experiment were shown an image and asked to discriminate between different anger classes. (31) has developed a novel approach for smile detection for faces in the wild which was optimized by training machine learning algorithms for improved pattern recognition using AU6. The architecture of AFFDEX.2.0 3.5, highlights that a combination of Action units is used to determine the evidence for different emotions(8). (25) suggested that action units 4,5,7,23 are dominant during the expressions of anger. (32) presented a comparative study between their research and darwin's observations. Darwin observed the presence of action units 4,5,24,38 for anger expressions while their research found either the presence of action units 4,5 or action units 7+22+23+24 during anger expressions. For compound expressions, the difference in the agreement of dominant action units were significant. For embarrassment, (33) highlighted that action units 12,24,51,54,64 were activated while (34) found evidence for activation of action units 7,12,15,52. These works present a basic premise that there are certain dominant action units associated with a specific emotion expression, however, agreement of dominant AU differs.

The important reasons behind these differences are mainly attributed to the inherent variations on how emotions are depicted and perceived; the cultural aspect of the emotions and the influence of gender, race, and ethnicity.

The other important aspect of difference in affect recognition is the nature of facial expression. Facial expressions can be posed or spontaneous. Most of the work in the literature involves posed facial expressions. Posed facial expressions may not reflect real-world naturalistic scenarios. (35) highlighted that there is a significant difference with respect to morphology

Table 2.1: Description of Action Units based on FACS (1; 2); rearranged based on the facial anatomy

Action Unit	Facial Muscles	Action Unit	Facial Muscles
1	Inner Brow Raiser	17	Chin Raiser
2	Outer Brow Raiser	14	Dimpler
4	Brow Lowerer	10	Upper-Lip Raiser
5	Upper-Lid Raiser	12	Lip-Corner Puller
7	Lid tightener	15	Lip Corner Depressor
43	Eyes	16	Lower Lip Depressor
41	Lid Droop	18	Lip Puckerer
42	Slit	20	Lip strechter
44	Squint	22	Lip Funneler
45	Blink	23	Lip Tightener
46	Wink	24	Lip Pressor
9	Nose Wrinkle	25	Lips part Depressor
11	Nasalabial Deepener	28	Lip Suck
6	Cheek Raiser	27	Mouth Strech
13	Cheeks Puffer	26	Jaw drop

and dynamic aspects between posed and spontaneous facial expressions. Posed expressions may be a result of voluntary facial movement which may be exaggerated and may reflect the actor's perception of emotion. (16) and (36) characterized that spontaneous expressions are genuine expressions and concealed emotions will be revealed; for which Ekman coined the term 'emotional leakage'.

FACS was considered as a promising approach and proved to be the basis for many related works in the domain of emotion recognition. (37) involved human FACS coders to differentiate between genuine and fake facial activity during pain. Similarly, in an interesting work (38), human FACS coders were used to identify suicidal signals in clinically depressed patients. However, in the year 1999, MS Bartlett highlighted the need for an automatic facial action coding system. They presented an argument that the time taken by human FACS coders to get trained and code a videotape are significantly high and an automated FACS would facilitate increased facial action coding speed with improved reliability and precision (39). This led to acquiring a patent on an automated facial action coding system, and the pipeline is presented in figure 2.3. There are multiple modern automation emotion recognition systems developed and such modern systems are highly advanced in terms of pattern recognition and FACS provides the basis for many such modern systems. Emotient and Affectiva are two such systems which was developed with the foundation of FACS (7; 24). Such systems integrate the foundation provided by FACS and utilize optimized pattern recognition algorithms to capture the smallest changes in the facial muscles to infer the implied emotion.

Emotient FACET is developed by Emotient and is now owned by Apple. It is based on the Computer Expression Recognition Toolbox (CERT) (7). The developers of FACET trained the algorithm on multiple databases of posed and spontaneous facial expressions. The underlying algorithm was evaluated against the Extended Cohn-Kanade dataset (40) and M3 database

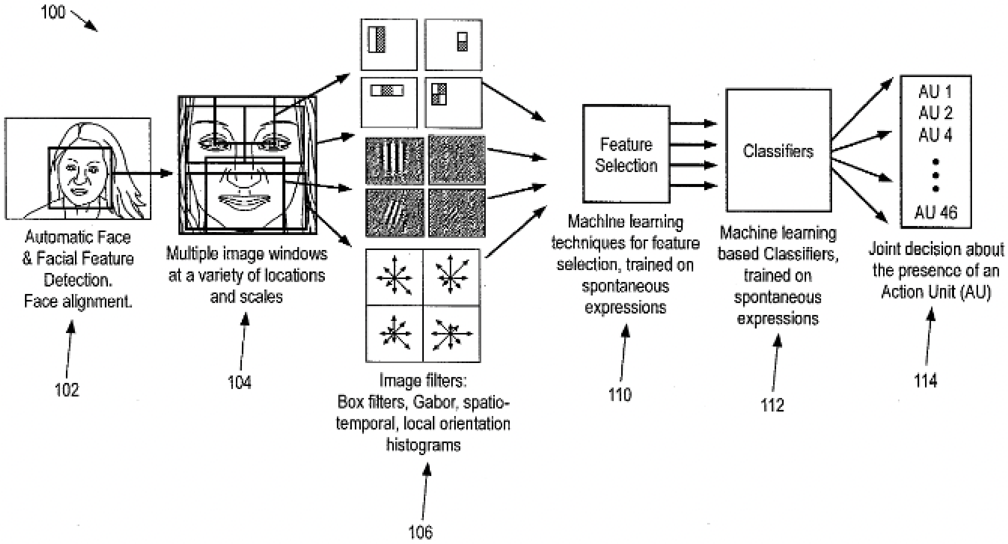

Figure 2.3: Pipeline for Automated FACS, figure from (5)

(41). Affectiva is an organisation by itself and it has developed a software development kit that facilitates facial expression analysis (24; 8). AFFDEX is a specialised toolkit that enables facial expression analysis in the wild. AFFDEX can detect the facial action units from the input video frame and recognise the presence of basic emotions. AFFDEX also estimates the Euler angles to identify the head orientation in a three dimensional space. The developers of the Affectiva have crowd sourced the dataset with naturalistic and spontaneous facial expressions and used active learning to track the face and extract HOG features. The comparison between the systems is presented in the table 2.2.

The performance of such automatic emotion recognition systems is evaluated against different sets of databases. Multiple databases like BP4D (42) are developed to facilitate research on posed facial expressions and other databases like (43) are developed with spontaneous or near spontaneous facial expressions. (44), compared the performance of Emotient, Affectiva, and AZURE against both posed and spontaneous facial expressions databases. Dupre. et all, observed that the performance of human observers was better than the automatic emotion recognition systems. However, within the tested systems, he observed that AZURE outperformed both Emotient and Affectiva. The systems had a better recognition accuracy of emotions from posed facial expressions rather than in a naturalistic setting. Similarly, the results from the studies conducted by (45) were in agreement with the results of (44); the recognition accuracy of Emotient and Affectiva were lower in spontaneous facial expressions. Further, they also observed that the performance of Emotient was better than Affectiva. (46) also observed that the recognition accuracy of Emotient for certain emotions was better than humans. The precursor study of the presented research compared the performance of different AERs across gender, ethinicity and age dimensions (47). However, the presented research adds profession of individuals as a new dimension of comparison between systems.

Table 2.2: Comparison of Emotient vs Affectiva

<table>
<tr><td colspan="2"></td><td>Emotient</td><td>Affectiva</td></tr>
<tr><td rowspan="3">Data</td><td>Training Database Type</td><td>Posed Facial Expressions</td><td>Spontaneous - 'Images in the Wild' Crowdsourced</td></tr>
<tr><td>Dataset Size</td><td>Unknown</td><td>1.8 Million</td></tr>
<tr><td>Validation Database</td><td>Extended Cohn-Kanade (Posed), M3 Dataset (Spontaneous)</td><td>Crowd-sourced dataset</td></tr>
<tr><td rowspan="3">Facial Expressions</td><td>Facial Landmarks</td><td>Six Facial Landmarks</td><td>34 Facial Landmarks</td></tr>
<tr><td>Action Units/ Facial Muscles</td><td>20 Action Units</td><td>18 Facial Muscles</td></tr>
<tr><td>Emotion Evidence</td><td>7 Primary Emotions</td><td>7 Primary Emotions</td></tr>
<tr><td colspan="2" rowspan="2">Head Orientation</td><td>Projection of Head in 3-dimensional space in terms of Euler angles (YAW, PITCH, ROLL)</td><td>Projection of Head in 3-dimensional space in terms of Euler angles (YAW, PITCH, ROLL)</td></tr>
<tr><td>Camera Perspective</td><td>Subject Perspective</td></tr>
</table>

2.2 Head Orientation

Head Pose Estimation Humans have the ability to quickly understand head orientation and head movements, and thereby possibly infer the non-verbal emotional cues with better accuracy. However, this is a challenging task in the field of computer vision. (48) describes head pose estimation as the "process of inferring the orientation of a human head from digital imagery" and summarises the methods used in head pose estimation and the various challenges associated with the task. The systems that are capable of estimating the head pose must also be invariant to the various imaging conditions. The head pose is either estimated as a discrete head direction or a head orientation in a 3-Dimensional space. (49; 50) employ novel approaches to estimate the head direction. (49) proposed a novel appearance-based method for estimating the head directions from individual scenes while (50) proposed a multi-spectral head direction estimation technique. (51) developed a fast and reliable head pose estimator based on random forests which were able to achieve state-of-art performance. The head pose is estimated from a static 3D image in terms of Euler angles. The proposed algorithm considers the head pose as a regression problem and utilizes an enormous dataset to optimize the predictions from the random forest. Similarly, (52) presented a robust algorithm to estimate the head pose of drivers in two degrees of freedom. Their approach involved using LGO histogram analysis followed by a support vector regression.

This indicates the basic premise that the scholars consider the head pose estimation problem in computer vision as a regression problem. In the first case, the 3D static image is used and in the later work, a prominent frame from the video stream is identified to estimate the head pose. The head poses in terms of Euler angles are estimated with reference to the global coordinate system or with reference to the camera coordinates. However, estimating the head pose in a continuous video is a tricky problem in computer vision.

(6) proposed a discriminative approach for frame-by-frame head pose detection. The pipeline of the head pose tacking from the presented work is illustrated in figure 2.4.

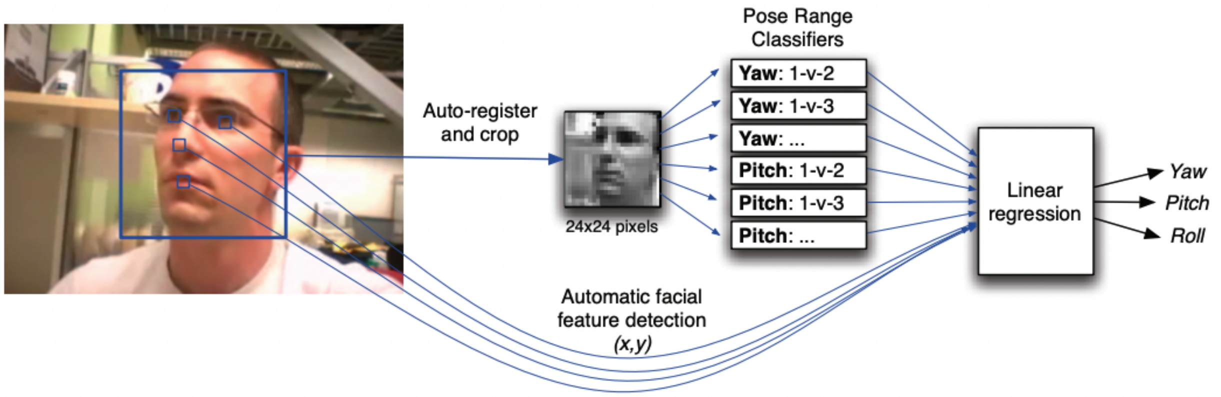

Figure 2.4: Pipeline for frame-by-frame head pose estimation, figure from (6)

Novel frame-by-frame by head pose tacker from (6) was trained and evaluated against the GENKI dataset and achieved an accuracy of 5.82○, 5.65○, and 2.96○ root-mean-square (RMS) error for each of the Euler angles. This algorithm was the basis of the head pose estimator used in Emotient. The head pose estimator of Emotient was adopted from this approach with custom optimizations (7). Emotient uses two pose range classifiers to estimate the head orientation. These classifiers produce the log probability ratio of the face in a particular angle range, which is then channelized as an input to the linear regressor to get the real values of Euler angles.

However, the architecture of Affectiva is different and it uses a different CNN-based pipeline to estimate the head orientation. (8) developed a custom facial landmark detector that detects the outer eye corners, nose tip, and chin. The output from the landmark detector is channelized through the convolutional neural networks to estimate the 3D head pose. These variations in the head pose estimation pipelines of AER indicate that, there is a significant difference in the architecture. This presents a scope for a validation study to compare the difference in the estimated orientation of the head and understand if the task at hand is invariant of the automatic emotion recognition system.

Head Orientation and Emotions

It was evident from Darwin's observations that emotional expression is not uni-modal, however, the impact and importance of different modalities are heavily researched. (34) presented the argument that emotional expressions depend on dynamic patterns of behavior ranging from facial actions to head movement, including gaze and scent. The consideration of different modalities has facilitated an improved understanding of how emotions are depicted and perceived in humans. (53; 54) conducted experiments to understand the body configurations which enabled better perception of pride expressions. They presented the results which indicated that the perception of pride expressions was higher when other aspects like, head orientation, body posture, and arm positions were considered along with facial expressions. (55) hypothesized that "different emotions are most effectively conveyed through specific, nonverbal channels of communication". They conducted experiments to present the evidence that, embarrassment was effectively communicated through eye gaze and head movements. These works illustrate that, the different channels of non-verbal communication like, head movements and gestures reinforce emotions.

(3) explored the relation between different body postures and the emotions. They conducted an experiment where 2 human annotators analysed at 224 videos for different posed emotions to understand the correlation of various body movements. From the results, (table **??**), we observe that, the authors were able to associate different head orientation for different emotions, and the head orientation is different for variation of the same emotion (Elated joy and happiness). These experiments identified evidence for the existence of emotion-specific head and body movements where, expressions for sadness were much more influenced by head movements than anger. The study concluded that there is a varying amount of influence of head movement in the display of different emotional expressions.

Table 2.3: Results from (3)

Emotion	**Head Orientation**
Elated Joy	Backward
Happiness	Downward
Shame	Downward
Pride	Backward
Boredom	Backward

In an interesting study (56), the impact of non-verbal cues and the interaction between the cues are investigated in the depiction of dominance and strength. The results indicate that a particular configuration of head orientation (raised head or bowed head + direct gaze) intensified the depiction of dominance and strength.

Similarly, there are many psychological experiments conducted to explore the influence of body movements on the personality traits of the individual. (57) presented the experiments where, the body movements of politicians were mapped to the movement of an animated stick figure to understand the relation between the body movements and the personality types like, consciousness and emotional stability of the individual. The experiment provided evidence that, the individual displaying low conscientiousness during the speech seemed to have an increased head movement.

(58) conducted experiments to understand the impact of head orientation and head movements by exploring the variations in Euler angles. The study utilized the BP4D database which contains videos of spontaneous facial expressions and analyzed the variation in the means of the Euler angles. They provided statistically significant evidence that, the mean for yaw and roll was close to zero for all the discrete emotions. This work indicates that, there may be a correlation between the change in Euler angles and the depiction of emotion. An interesting study (59) was also presented where the authors tried to make a 'humanoid' robot express different emotions. The emotions depicted by the robots were perceived by humans and the

authors presented evidence of the impact of head movements on the perception of emotions. This study is very peculiar and interesting since the robot does not have any facial muscles and the head movements are understood in isolation.

Further, there are various studies (60; 61; 62) where, deep neural networks are employed to detect emotion continuously from videos. (60) have explored the impact of head pose and gaze in continuous emotion recognition task and presented a novel framework to fuse non-verbal cues with facial expression to detect emotions. All these studies indicate a strong correlation between head orientation and the depiction of emotion.

2.3 Summary

The potential summary of the studies reviewed so far is tabulated (table 2.4 on the basis of whether humans or machines were used for depiction and perception of emotions. The emotions can be displayed and perceived by both humans and machines. The displayer of the emotions is presented as the vertical axis and the perceiver of the emotions is presented as the horizontal axis. Only in the case, (57) and (59), machines were used to display emotions. The core similarity for most of the reviewed works is that, the ground truth for emotion is manually coded and not automatically recognized. This presented a scope for a validation study to understand the impact of head orientation when automatic emotion recognition systems were used to perceive emotions. In the presented work, the following research questions are explored:

1. Is the estimated emotion evidence from semi-spontaneous videos by different automatic emotion recognition systems invariant of the individual's profession?
2. Is the computation of head orientation independent of the facial recognition systems used?
3. What is the relationship between the head orientation and the depicted emotion?
   (a) Is there a relationship between head direction and emotion?
   (b) Is there an association of a specific head orientation for each emotion?

Table 2.4: The matrix represents the reviewed studies grouped on the basis of displayer and perceiver of emotions presented in a chronological order, as discussed with the supervisor

		Perception of Emotions	
		Humans	**Machines**
Display of Emotions	**Humans**	Wallbott (1998) Tracy and Robins (2004,2007) App et al. (2011) Toscano et al(2018) Keltner et al (2019)	Dael et al. (2012) Li et al. (2017) Wu et al. (2019) Ahmad et al. (2021) Presented Work
	Machines	Koppensteiner and Grammer (2010) Johnson and Cuijpers (2019)	Not known to the author

The following chapter describes the curated dataset and methodology designed to investigate these research questions. Two automatic emotion recognition systems (Emotient and Affectiva) are selected to evaluate different hypotheses. The experiments and results are presented in the subsequent chapters.

Chapter 3

Design Methodology

The proposed methodology pipeline to understand the impact of the head pose in automatic emotion recognition systems is designed in 5 phases. This section presents the overview of the methodology pipeline and then discusses each phase in detail. The overview of the pipeline is illustrated in the figure 3.1.

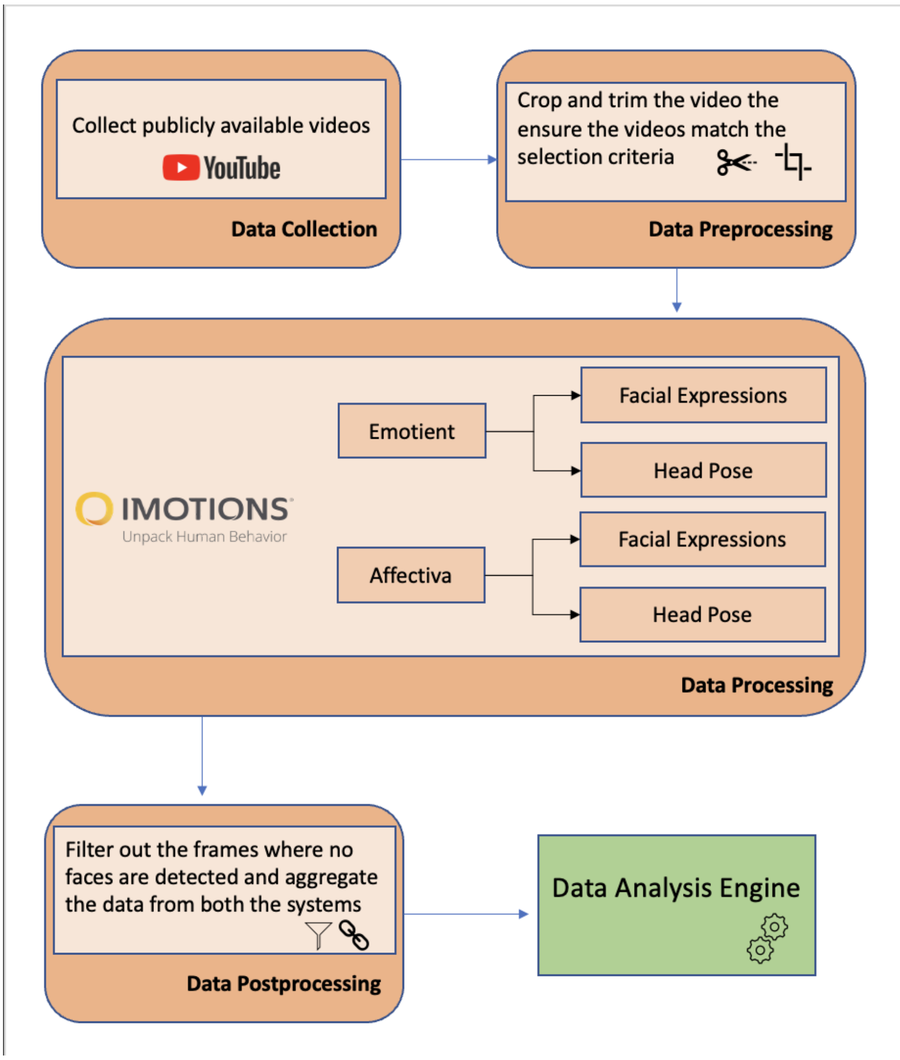

Figure 3.1: Overall Workflow Pipeline: Data Collection, Data Preprocessing, Data Processing, Data Postprocessing, Data Analysis

3.1 Overview of Approach

The initial phase is data collection and profiling, where the publicly available videos are collected from YouTube. This raw data is channelized through a series of pre-processing steps to ensure that the downloaded videos meet the selection criteria. Two automatic emotion recognition systems (Emotient and Affectiva) are used to recognize facial expressions and estimate head pose. These systems use state-of-the-art algorithms under the hood to identify the facial landmarks, calculate the activation of action units(or facial muscles) and estimate the head pose. The pre-processed video is broken down into discrete frames and is processed by these systems. The processed frames where no faces are detected are discarded. The frames are further filtered by performing an 'inner join' operation to ensure only the frames processed by both the systems are selected and the remaining frames are filtered out. The aggregated data from both systems are sent to the data analysis phase for further analysis. Each phase is explained in detail below.

3.2 Dataset

The profession of the individuals is set as the primary selection criteria for the dataset. CEOs, politicians, and spokespersons are charismatic people who are in the position to make decisions and may share common personality traits. The videos of these individuals addressing the mass media by either giving a speech or an interview on a public news channel are selected. This ensures controlled lighting conditions with minimal distractions. These individuals are prepared to give a speech (they may display posed expressions) but also answers impromptu questions (they may involuntarily exhibit their idiosyncrasies,i.e, spontaneous expressions), and thus the curated dataset is semi-posed or semi-spontaneous. The videos are collected from popular

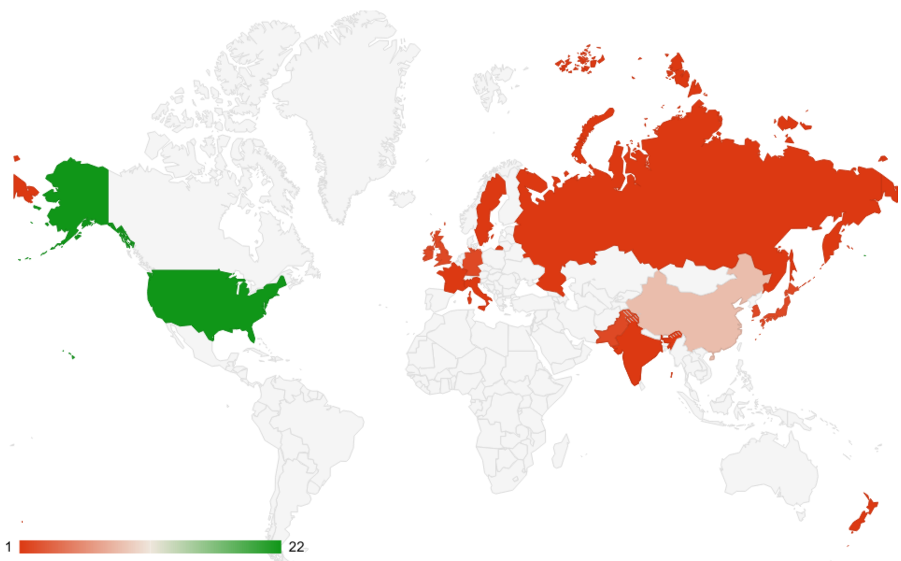

Figure 3.2: Demographic Distribution of selected individuals

news channels and public media channels. This ensures that the videos are shot by leading videographers and thus the light, noise, and other distractions are well controlled. A total of 221 videos of Politicians, CEOs, and Spokespersons are collected from various YouTube channels. These videos correspond to 65 unique individuals from different demographics. The total run time of videos is more than 14.5 hours. An open-source YouTube downloader is used to batch download videos. The demographic distribution of the data is presented below (figure 3.2)and summarised with respect to gender in the table 3.1.

Table 3.1: Data demographic profile: by nationality and gender, the count of individuals, and the total number of videos

	Individuals		**Videos**	
Nationality	Female	Male	Female	Male
China	3	7	14	23
France	1	0	2	0
Germany	1	1	5	5
India	1	7	2	18
Ireland	5	7	28	28
Italy	0	1	0	2
Japan	0	1	0	1
New Zealand	1	0	5	0
Pakistan	0	2	0	6
South Korea	0	2	0	4
United Kingdom	0	1	0	5
United States	12	12	41	32

The subjects are diversified in terms of race, gender, and occupation. The distribution of subjects based on age and profession is visualised in (figure 3.3) and summarised in the table 3.2. The youngest individual is 28 years old and the oldest individual is 94 years old with 20% of the individuals in the range of 61-72 years.

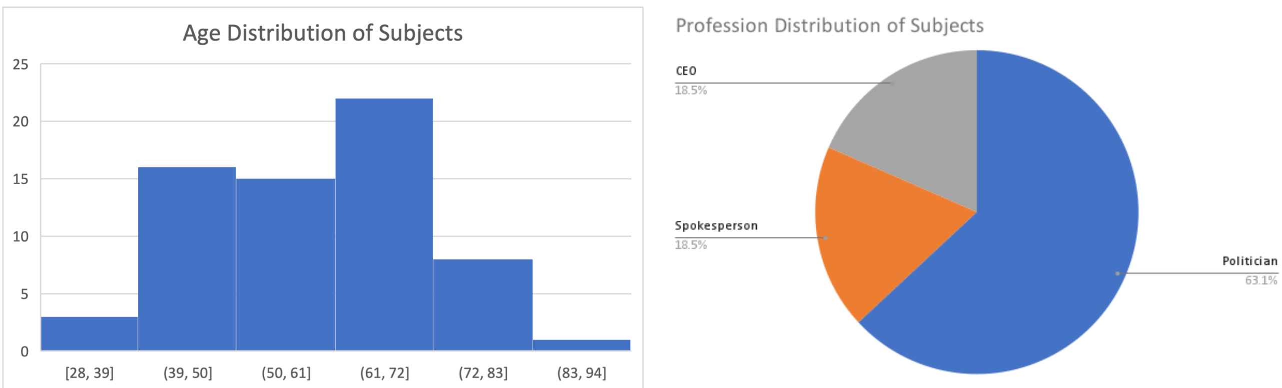

Figure 3.3: Age and Occupation Distribution of Subjects

The curated dataset is channelized through a series of pre-processing steps where the downloaded video is cropped and trimmed to ensure the presence of only one person in any given frame. Additionally, in the frames where no faces are detected, are filtered out. Pre-processed videos are then given as input to the emotion recognition systems which break them into dis-

Nationality	Age range	Occupation CEO	Politician	Spokesperson
China	49 – 73	1	6	3
France	49 – 49	1	0	0
Germany	62 – 67	0	1	1
India	28 – 81	3	5	0
Ireland	36 – 81	1	9	2
Italy	46 – 46	1	0	0
Japan	73 – 73	0	1	0
New Zealand	41 – 41	0	1	0
Pakistan	46 – 69	0	1	1
South Korea	69 – 71	0	2	0
United Kingdom	57 – 57	0	1	0
United States	39 – 91	5	14	5

Table 3.2: Age ranges and counts of individuals in each occupation, by nationality

crete frames for further processing. The architecture and workflow of the selected systems are explained in the following section.

3.3 Automatic Emotion Recognition Systems

The crux of the presented research work is to validate whether the existing automatic emotion recognition systems are invariant across different attributes when they capture facial expressions and estimate head poses from videos. Two automatic emotion recognition software systems (Emotient and Affectiva) are considered for comparative analysis. Both the systems are available via the iMotions software suite and access to iMotions is facilitated by the educational license granted to Trinity College Dublin.

3.3.1 Emotient FACET

Emotient FACET analyzes the input frames from the video for various facial expressions to detect the presence of seven primary emotions (joy, fear, anger, sadness, contempt, disgust, and surprise) (7). Emotient FACET identifies six facial landmarks and captures the evidence for seven primary emotions. Emotient FACET produces evidence scores for the mentioned emotions and the underlying action units as per the FACS. The pipeline involved in the calculation of emotion evidence by Emotient is presented in the figure 3.4. These evidence scores represent the log-likelihood of the presence of emotion in the given frame. The evidence score can either be negative or positive and indicates the probability of a particular emotion in a given frame. For instance, negative anger evidence indicates that there is less than 50% probability of the individual expressing anger in the given frame. If the emotion evidence is zero, there is a 50% chance for the presence of that emotion in the given frame.

Emotient FACET uses a discriminative frame-by-frame approach to estimate the head orientation. For each frame, the system identifies the bounding box for the face and registers the

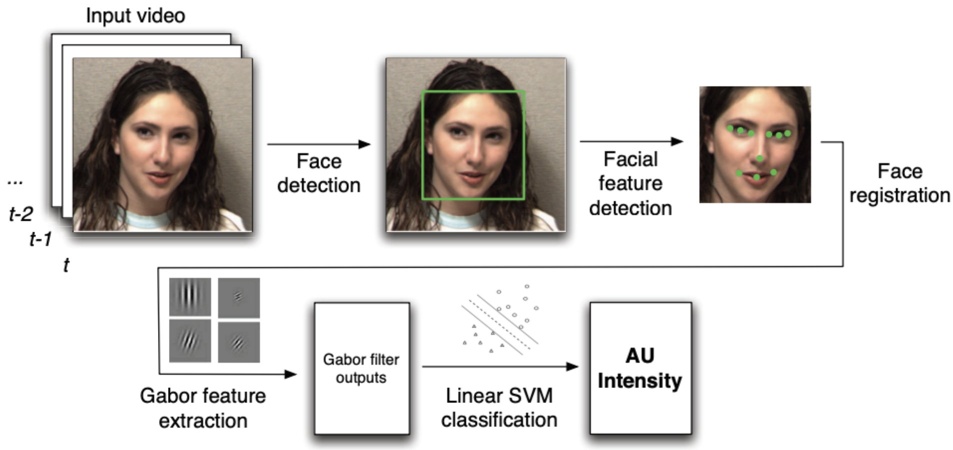

Figure 3.4: Emotient Pipeline involved in the calculation of AU evidence, figure from (7)

faces. The pixel values from the bounding box are given to an array of pose range classifiers which are then channelized to a linear regression model to get the real values of Euler angles.

3.3.2 Affectiva AFFDEX

Affectiva is also an automatic emotion recognition software similar to Emotient. The pipeline followed by Affectiva is illustrated in the figure 3.5.

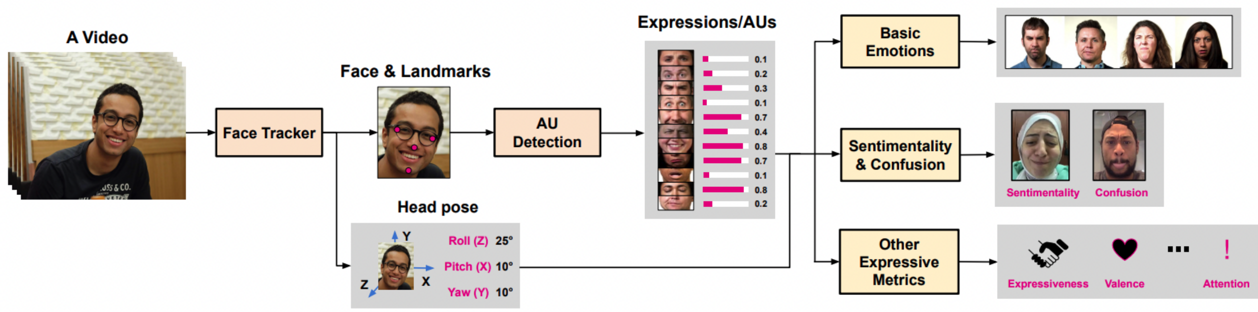

Figure 3.5: Pipeline of Affdex from (8)

AFFDEX uses multiple support vector machine classifiers along with kernel estimation to generate a score for each of the identified facial action units. Affectiva employs the same strategy to generate a score for the expression of emotions since they are a linear combination of facial action units.

The scoring scheme of Affectiva ranges from 0 to 100 where 0 indicates the absence of the action unit and 100 indicated the presence of the action units.

3.3.3 Data Processing in Emotient and Affectiva

Both these systems take a video as an input, break it into discrete frames, and provide the output with reference to facial expressions for every frame. On average, the videos are processed at 33 to 40 frames per second. If a video duration is 5 minutes with 40 fps, the systems are expected to produce an output of 12000 frames. Both systems produce the activation of action units (or facial muscles), evidence for emotion, and head pose (in terms of Euler angles) for each of these 12000 frames.

In this research work, the curated dataset has 221 videos with a total duration of 14.5 hours. During the post-processing, the frames where no faces are identified by both systems are discarded. Before pre-processing, the overall dataset has over 1.33 million frames processed by Emotient and 1.3 million frames processed by Affectiva. For further processing, frames that are processed by both systems need to be identified. Thus, the frames from Emotient and Affectiva are merged using an 'inner join' operation using the name and media time attributes to create an aggregated dataset of 1278911 frames. 34056 frames from Affectiva and 56772 frames from Emotient are discarded. (Data Retention = 97.4% frames from Affectiva, 95.7% frames from Emotient). The frames processed with respect to videos from each nationality are summarised in the table 3.3. We can clearly observe that Emotient can process more frames for a given video as compared to Affectiva.

Table 3.3: Number of frames processed by each system by nationality for all the videos in the dataset.

	#videos	Total Frames in Affectiva	Total Frames in Emotient	Common Frames in Both Systems
China	37	202025	206140	200979
France	2	7995	7364	7354
Germany	10	40602	40796	40445
India	17	88060	112672	85780
India	3	11758	10943	10928
Ireland	56	287728	283672	282864
Italy	2	8356	8370	8356
Japan	1	20229	20213	20206
New Zealand	5	25128	25750	25002
Pakistan	6	72585	59553	58965
South Korea	4	35184	35237	35172
United Kingdom	5	47068	52387	46470
United States	73	466249	472586	456390
Total	**221**	**1312967**	**1335683**	**1278911**

Each of these frames has information about the emotion evidence and the estimated head orientation. Additional attributes are computed from these attributes and the flow of analysis is explained in the following section.

3.3.4 Facial Expressions

Both systems produce an output file for each video processed which contains data for every frame. For each frame, both systems produce the evidence captured for 6 basic emotions along with the evidence of action units (in case of Emotient) or facial muscles (in case of Affectiva). For demonstrating the key aspects of the Emotient and Affectiva, I have taken a video of a prominent and well-known CEO of a large multinational, and the output from each system is explored in detail below.

Emotient FACET

For this particular video, Emotient produces 7513 frames. Figure 3.6 illustrates the change in emotional evidence over time.

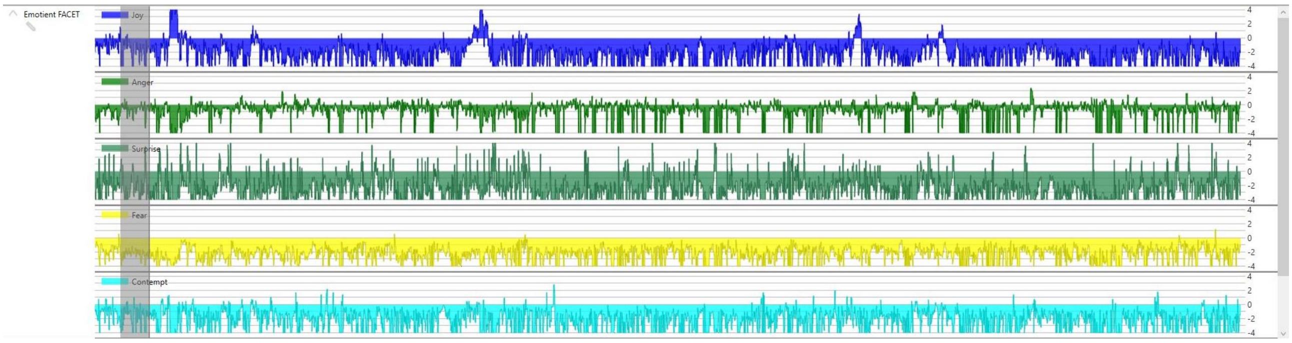

Figure 3.6: Processed OUTPUT from Emotient

For each frame, Emotient produces the estimated evidence for 20 action units and 6 basic emotions. The evidence scores represent the log-likelihood of the presence of emotion or activation of an action unit. These log-likelihood values can be converted to intensity values ((63)) which are within the range of 0 to 1 using the formula mentioned below. For each frame at time t,

$$Intensity(emotion, t) = 10^{evidence(emotion,t)}/[1 + 10^{evidence(emotion,t)}]$$

$$Intensity(actionunit, t) = 10^{evidence(actionunit,t)}/[1 + 10^{evidence(actionunit,t)}]$$

The emotion intensity values are computed from the corresponding evidence values for the selected video and the comparison is presented in the figure 3.7. The same formula can also be used to calculate the intensity of the action units for each frame. These intensity values can be further used to deduce the probable emotion displayed by the subject at time t. For each frame at time t, the emotion with the highest intensity value is considered the probable emotion at time t. The interaction of the action units and emotion intensities with the variation of Euler angles is explored in the next section.

Affectiva AFFDEX

For the same video considered above, Affectiva produces 7948 frames. Figure 3.9 illustrates the change in emotional evidence over time. For each frame, Affectiva produces the estimated evidence for 19 facial muscles and 6 basic emotions. Affectiva estimates the activation of the following facial muscles: brow furrow, brow raise, lip corner depressor, inner brow raise, eye closure, nose wrinkle, upper lip raise, lip suck, lip press, mouth open, chin raise, smirk, lip pucker, cheek raise, dimpler, eye widen, lid tighten, lip stretch, jaw drop.

The facial muscles are similar to the action units produced by Emotient. The mapping between the facial muscles and action units is presented in the table 2.1. The evidence scores

	anger_evidence	surprise_evidence	fear_evidence	disgust_evidence	sadness_evidence
950159	-0.156431	-1.737062	-0.844534	-0.631596	-1.351111
950160	-0.141268	-1.464252	-0.727871	0.152171	-0.227760
950161	0.400079	-2.150489	-0.533256	-0.062473	-0.061143
950162	0.089757	-3.444975	-0.973821	-0.146615	-0.170066
950163	0.403919	-2.348254	-0.909369	-0.647776	-1.211179

	anger_intensity	surprise_intensity	fear_intensity	disgust_intensity	sadness_intensity	probable_emotion
950159	0.410912	0.017991	0.125142	0.189340	0.042654	anger
950160	0.419389	0.033196	0.157628	0.586712	0.371815	disgust
950161	0.715290	0.007022	0.226555	0.464099	0.464861	anger
950162	0.551485	0.000359	0.096015	0.416394	0.403334	anger
950163	0.717087	0.004465	0.109691	0.183688	0.057930	anger

Figure 3.7: Emotion evidence vs Emotion intensity values. Emotion Intensity values (2nd table,DOWN) are computed from evidence values (1st table,UP) for each frame in Emotient

	950159	950160	950161	950162	950163
au1_evidence	-0.247725	0.045125	0.329375	0.011866	0.328772
au2_evidence	-1.718970	-1.115824	-1.321999	-2.171031	-1.660116
au4_evidence	2.068842	0.782621	1.632690	2.616167	2.013071
au5_evidence	-0.638071	-1.001014	-0.664970	-0.367654	-0.239733
au6_evidence	0.314163	-1.089628	-0.696296	0.246293	-0.623834
au7_evidence	1.178233	-1.185795	-0.050418	0.899249	0.356886
au9_evidence	-0.555678	-0.370985	0.063678	0.949017	0.394269
au10_evidence	0.726597	-0.119503	0.346232	0.969439	0.615952
au12_evidence	-0.685854	-0.725389	-0.831245	-1.116009	-1.014934
au14_evidence	-0.917325	-0.102672	0.007190	-0.332801	-0.400524
au15_evidence	-1.073626	-0.184437	0.421907	0.107468	-0.336692
au17_evidence	-1.371190	-0.252870	0.303045	0.219007	-0.569803
au18_evidence	-0.110388	0.423801	1.216285	-0.298166	-0.138558
au20_evidence	0.447264	-0.126657	0.040152	0.334607	0.236949
au23_evidence	-0.810891	-0.716663	-0.214989	-0.282521	-0.181021
au24_evidence	-1.184146	0.177763	0.656084	0.264166	0.168596
au25_evidence	1.694957	-1.375507	-0.895010	-0.505097	0.222686
au26_evidence	0.384197	-1.374362	-0.812292	-1.134490	-0.208877
au28_evidence	-2.051750	-0.576434	-0.539418	-0.371022	-0.345008
au43_evidence	0.643013	2.202598	1.129011	0.191710	0.475893

Figure 3.8: Evidence for the activation of each action unit for the same reference frames in figure 3.7. We can observe that in each frame, Emotient is able to provide activation evidence for more than one action unit

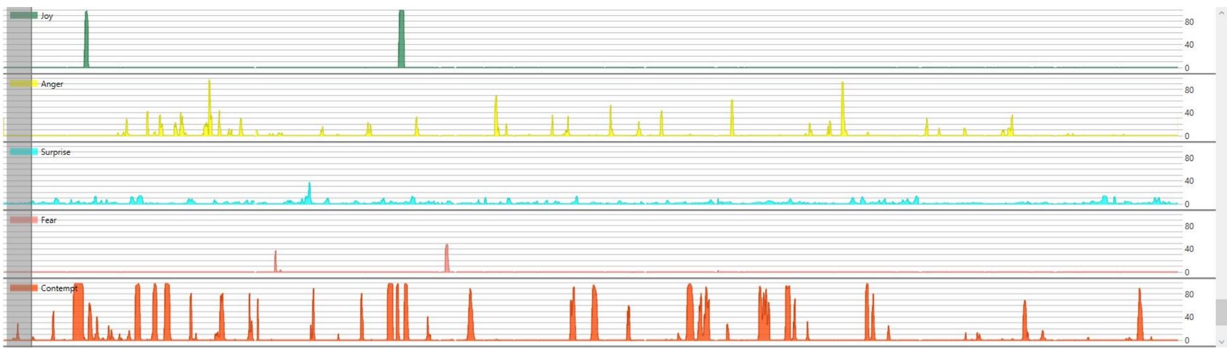

Figure 3.9: Processed Output from Affectiva

are in the range of 0 to 100, where a score of 0 indicates no evidence of emotion and 100 indicates concrete evidence for emotion. The emotion evidence scores and evidence of activation of facial muscles in compared in the figure 3.10. For each frame at time t, the emotion with the highest evidence value is considered as the probable emotion at time t.

	joy	anger	surprise	fear	disgust	sadness	probable_emotion
933803	99.92060	0.000074	2.290420	0.000075	0.000365	9.770000e-08	joy
933804	99.92014	0.000073	2.254931	0.000074	0.000209	8.290000e-08	joy
933805	99.91979	0.000075	2.302038	0.000075	0.000594	1.150000e-07	joy
933806	99.91876	0.000076	2.307969	0.000075	0.001252	1.480000e-07	joy
933807	99.91516	0.000104	2.132796	0.000075	0.000154	8.270000e-08	joy

Figure 3.10: Evidence for each emotion values ranges from 0-100. We can observe that Affectiva produces high evidence values for dominant emotion('joy' in this case) and very low values for other emotions

	933803	933804	933805	933806	933807
brow_furrow	2.207062e-03	1.860494e-03	2.568433e-03	3.183842e-03	1.716937e-03
brow_raise	4.020000e-05	5.900000e-05	2.460000e-05	1.590000e-05	7.790000e-05
lip_corner_depressor	2.680000e-12	2.330000e-12	3.230000e-12	3.460000e-12	2.320000e-12
innerbrowraise	4.134322e-01	4.528391e-01	4.178778e-01	3.868108e-01	5.674036e-01
eyeclosure	4.380000e-05	1.830000e-05	1.254610e-04	4.212010e-04	2.170000e-05
nosewrinkle	1.275790e-02	7.624209e-03	1.624877e-02	1.383948e-02	3.669962e-03
upperlipraise	6.146240e-04	5.236660e-04	6.003600e-04	3.505680e-04	3.553190e-04
lipsuck	5.050961e+01	6.168930e+01	4.064729e+01	2.541039e+01	6.748344e+01
lippress	9.997277e+01	9.996986e+01	9.997614e+01	9.996255e+01	9.995589e+01
mouthopen	9.952147e+01	9.885677e+01	9.971713e+01	9.983777e+01	9.649173e+01
chinraise	3.153859e+00	4.790359e+00	2.441139e+00	1.811554e+00	9.364194e+00
smirk	7.136990e-01	6.922117e-01	7.675843e-01	8.095307e-01	7.946619e-01
lippucker	1.962419e-01	2.327270e-01	1.771755e-01	1.337082e-01	2.737641e-01
cheek_raise	2.289753e+01	2.356906e+01	2.138368e+01	2.012391e+01	2.168822e+01
dimpler	1.395012e+01	2.544570e+01	8.516842e+00	5.187264e+00	4.282820e+01
eye_widen	7.035170e-04	5.491190e-04	7.686020e-04	8.903130e-04	5.326080e-04
lid_tighten	7.402930e-01	7.876368e-01	6.972672e-01	6.913351e-01	8.153190e-01
lip_stretch	9.999847e+01	9.999864e+01	9.999819e+01	9.999763e+01	9.999875e+01
jaw_drop	2.767657e-02	1.919531e-02	3.867779e-02	5.244371e-02	1.373033e-02

Figure 3.11: Evidence for the activation of each facial muscles for the same reference frames in figure 3.10. We can observe that in each frame, Affectiva also produces activation evidence for more than one facial muscle

3.4 Head Pose Estimation

The head orientation is estimated by both systems in terms of Euler angles. Euler angles are measured in terms of radians and the pipeline followed by both systems to estimate the Euler angles is illustrated in the figure 2.4. Yaw angle captures the movement of the head on the horizontal axis. A negative yaw angle indicates that the head direction is to the left and a positive yaw angle indicates that the direction of the head is right. The pitch angle captures the head orientation in the vertical axis and the roll angles provide information about the head tilt. The estimated Euler angles vary over media time as the subject moves their head. The variation in the Euler angles from the video of a prominent politician from the United Kingdom is visualized as a time series and illustrated in the figure 3.12. Additionally, attributes like head direction, the relative change for each frame, and the angular velocity are derived from these Euler angles. The angular velocity in each direction is computed as the difference in the relative change of Euler angles of subsequent frames divided by the frame rate. The estimated Euler angles from both systems are compared in the figure 3.13.

Figure 3.12: Distribution of Euler Angles in both systems near common agreement index

Figure 3.13: Comparison of estimated Euler angles from both systems

	yaw_emotient	pitch_emotient	roll_emotient	yaw_affectiva	pitch_affectiva	roll_affectiva
0	9.23004	-0.824545	3.010365	15.52397	9.86270	39.726630
1	12.62409	-2.208452	1.953101	18.43913	14.09491	11.158320
2	11.87108	-2.238091	1.861500	19.64368	11.88924	12.161330
3	11.56567	-1.916042	1.491796	18.91169	12.77740	11.114400
4	10.92296	-1.818074	2.306007	17.20033	12.41543	9.598939

We can observe that both systems seem to have a disagreement in the estimated Euler angles. The variation in the Euler angles and their relation with emotion is explored and results are presented in the next chapter. The selection of statistical tests to test the various hypotheses and the rationale behind the selection is explained in the next section.

3.5 Statistical Processing and Hypothesis Testing

In the presented work, a battery of statistical tests is utilized to establish a statistical significance and test the various hypotheses. The rationale behind selecting these tests and the details of their implementation are presented below.

Shapiro-Wilk Test The Shapiro-Wilk test is used to check whether the data follows a normal distribution. We need to understand the underlying data distribution to select the set of statistical tests to evaluate the hypothesis built on this data. If the data follows a normal distribution, a set of parametric tests can be used to evaluate the hypothesis. Non-parametric tests are not built on any assumptions of data distribution and thus can be used if the data does not follow a normal distribution.

Shapiro-Wilk tests present the null hypothesis that the data was drawn from a normal distribution (64; 65). The Shapiro-Wilk tests return a W-statistic with a p-value. If the p-value is less than 0.05 we reject the null hypothesis that the data is drawn from a normal distribution. If the p-value is greater than 0.05, the null hypothesis can be accepted and we can infer that the data follows a normal distribution.

The rationale behind using this test is to check if the data produced from Emotient and Affectiva follows a normal distribution. The Shapiro-Wilk test is implemented using the 'scipy' package in python. The W-statistic and the associated p-Value are tabulated in the table below **??**. We can observe that the data for the estimated Euler angles from both systems do not follow a normal distribution.

Table 3.4: Results of Shapiro-Wilk test on the Estimated Euler angles

		W-statistic	**p-Value**
Emotient	**Yaw**	0.9941	0
	Pitch	0.997	0
	Roll	0.9962	0
Affectiva	**Yaw**	0.9959	0
	Pitch	0.9954	0
	Roll	0.9962	0

Kruskal-Wallis Rank sum test Kruskal-Wallis Rank sum test is used to check if there is a significant statistical difference between two or more groups. This test is non-parametric and therefore does not assume anything about the underlying data distribution. Kruskal tests rank the data to determine the difference. If the difference between both groups is small, the average rank of both groups will be similar with the same median. However, if there exists a strong statistical difference between the groups, there will be a different median and the average rank of the group will differ. Kruskal-Wallis tests present a null hypothesis that the population median of all the groups is equal. The tests produce a test statistic with a p-value (α). If α is ≤ 0.05, it indicates that the alternate hypothesis can be accepted, and there exists a significant difference between the groups. However, if α is $>$ than 0.05, it indicates that the null hypothesis can be accepted and there exists no statistical difference between the groups.

The rationale behind using these tests in the presented work is to understand the statistical difference between different groups of interest selected from the aggregated data between the systems. This test is also used to evaluate the difference between the systems(Emotient and Affectiva) on the estimation of the same physical correlates. (Emotion evidence and Head orientation).

Spearman's Rho Correlation Spearman Rank Order correlation is a non-parametric test that produces a rho coefficient with an associated p-value. This test can be used to measure the monotonicity of the relationship between two datasets (66). The rho value ranges from -1 to 1. The sign of the rho value indicates the direction of the relationship. If the value is greater than 0, it indicates a positive association between the variable, i.e, if one variable increases, the other variable also increases. If the rho coefficient has a value lesser than 0, it indicates an inverse relationship between the variables,i.e, if one variable increases, the other variable decreases. The magnitude of the rho value indicates the strength of the association. The strength of the association is classified as follows: Very weak (0.00 - 0.19), weak (0.20 - 0.39), moderate(0.40 - 0.59), strong (0.60 - 0.79) and very strong (0.80 - 1.00).

The spearman test presents a null hypothesis that the rho coefficient is 0 and the two data do not have any association. If the p-value for the produced rho value is less than 0.05, the null hypothesis can be rejected and we can infer that there is a strong association between the tested variables. The rationale behind using this test in the presented work is to understand the association between estimated Euler angles and emotion evidence produced by both these systems.

Cohen-Kappa Score In the presented work, multiple contingency tables are created to compare the head orientation estimates from both Emotient and Affectiva. To understand the agreement between the systems, the cohen-kappa score is evaluated and the variation of the cohen-kappa score for different emotions is calculated and compared. Cohen-kappa score represents the level of agreement between the two systems. The cohen kappa score is given by:

$$\kappa = (p_o - p_e)/(1 - p_e)$$

The null hypothesis here is that the value for κ is 0 and there is no agreement between the systems. If the p-value of the estimated score of kappa is less than 0.05, and the magnitude of the kappa score is high, it indicates that there is a high agreement between the systems and the alternative hypothesis can be accepted. If the p-value of the estimated score of kappa is greater than 0.05, it indicates that there is no agreement between the systems, and the null hypothesis is accepted.

3.6 Summary

In this section, the designed methodology pipeline is explained in detail. The features used from both systems to investigate the research questions are presented and compared. In the next section, the hypotheses of the presented work along with the experiments carried out for investigation and their results are detailed.

Chapter 4

Experiment and Results

4.1 Facial Expression

In this section, different experiments are conducted to understand the difference between the automatic emotion recognition systems. Further, multiple hypotheses are presented to investigate the first research question: "Is there any systematic and statistical difference between the automatic emotion recognition systems based on the individual's profession?"

4.1.1 Evaluation of agreement between Emotient and Affectiva

Both systems estimate the evidence for 6 basic emotions for every frame in the video. However, a system may be sensitive to a particular emotion. The system may be able to recognise the evidence for such emotion easily, than compared to others. To understand the sensitivity, the relative percentage of frames where, evidence for one particular emotion was higher than other emotions is evaluated and presented in figure 4.1.

	Overall	Politician	Spokesperson	CEO
joy	0.362719	0.370695	0.293166	0.494562
disgust	0.246453	0.208806	0.388520	0.070879
fear	0.177378	0.186391	0.122938	0.268418
surprise	0.129166	0.129464	0.145311	0.089126
anger	0.043927	0.055825	0.038114	0.005650
sadness	0.040358	0.048819	0.011951	0.071366

	Overall	Politician	Spokesperson	CEO
surprise	0.422556	0.342466	0.520802	0.534992
anger	0.281523	0.395214	0.130390	0.134769
joy	0.137417	0.103217	0.149994	0.217788
disgust	0.082600	0.095530	0.095850	0.032379
sadness	0.066600	0.048158	0.102258	0.078127
fear	0.009305	0.015414	0.000706	0.001946

Figure 4.1: Relative Percentage of Frames for each emotion by profession

From figure 4.1, we can observe that from the same set of videos, Emotient is able to recognise the highest evidence for joy in 36% of the frames while Affectiva is able to recognise joy in only 13.7% percent of the frames. The difference between the systems is further explored at the action unit level by comparing the frequency of action of different action units. Emotient provides the activation evidence for 19 action units while Affectiva provides evidence for 19 facial muscles. An activation threshold is set at 50%. So an action unit is considered to be activated if the evidence is greater than the activation threshold. The comparison of the frequency of activation in both systems is presented in the figure 4.2.

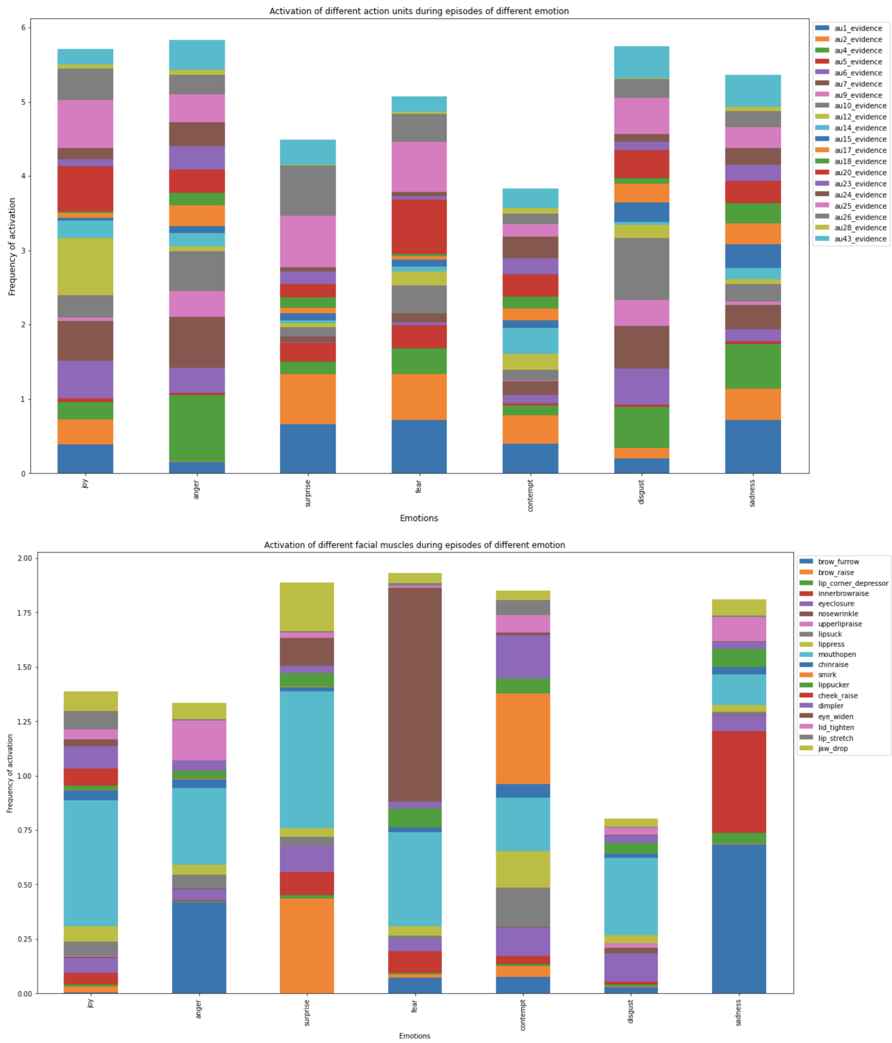

Figure 4.2: Frequency of activation of action units in Emotient (TOP), Frequency of activation of facial muscles in Affectiva (BOTTOM)

From the above figure, we can observe that there is a clear variation in the frequency of activation for different emotions. In Emotient, we can see a clear evidence for activation for action unit 12 for joy and action unit 4 for anger. For the same emotions, we can see a higher activation for mouth open and chin raise from Affectiva. Similarly, this observation also presents the scope to look at insignificant action units/facial muscles. For each emotion, certain action units or facial muscles do not cross the activation threshold even once for the entire dataset. Such action units and facial muscles are identified and tabulated (Table 4.1). This illustrates that there is a variation in the frequency of activation, however, the variation with respect to profession and the statistical significance of the difference is explored below.

The following analysis explores the difference in frequency of activation of action units and facial muscles with respect to the profession. The dataset is filtered based on profession and the number of frames where the activation of action units is greater than the activation threshold is computed. The results are visualised as a line plot and presented in the figure 4.3.

Table 4.1: Summary of action units and facial muscles whose frequency of activation is zero in the entire dataset

Emotion	**Insignificant Action Unit**	**Insignificant Facial Muscle**
Joy	AU {18}	Brow Furrow, Lip Puckerer
Anger	AU {2,5}	Cheek Raise
Surprise	AU {9,12,14}	Nose Wrinkle
Fear	AU {9,17,6,23}	Nose Wrinkle, Cheek Raise
Sadness	AU {12,6}	Cheek Raise, Brow Raise
Disgust	AU {18}	Cheek Raise Lip Puckerer

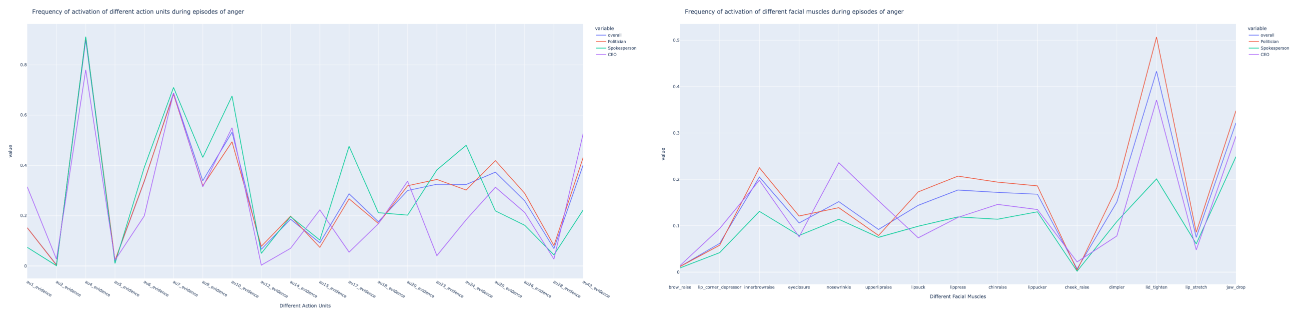

Figure 4.3: Frequency of activation of different action units in Emotient (LEFT) compared with the frequency of activation of different facial muscles in Affectiva (RIGHT) with reference to the profession when both systems detected evidence for anger.

(2) presented a comparative study of dominant Action units and facial muscles for different emotions, where AU4 and lid tighten were associated with anger. We can observe from the figure 4.3 that the frequency of action of significant AU seems to be the same across professions. For instance, there is a higher frequency of activation of AU4 for all three professions when anger was detected by Emotient, however, we can observe a variation in other action units. Similarly, in Affectiva, there is a higher activation for lid tighten across professions when the system detects evidence of anger, however, there is a variation in frequency for other muscles.

Statistical Significance

This section investigates if the similarities and differences observed above are statistically significant. The distribution of evidence from both systems across professions is compared at two levels: intra-system and inter-system. To compare the data distribution within the system, the data is filtered based on the one-vs-rest approach. All the frames corresponding to CEOs are grouped together and compared with all the frames corresponding to politicians and spokespersons (Not CEOs). To understand the difference in the data distribution, the Kruskal-Wallis rank sum test is utilised and the hypothesis is described below:

H_0***: The distribution of*** e_i ***in*** AER_i ***for*** p_i ***and*** $p_i^{'}$ ***is same***
H_A***: The distribution of*** e_i ***in*** AER_i ***for*** p_i ***and*** $p_i^{'}$ ***is not same***,where
$e_i \in joy, anger, surprise, fear, disgust, sadness,\ AER_i \in Emotient, Affectiva,$
$p_i \in CEO, Politician, Spokesperson$

The kruskal test is computed and based on the pValue either null hypothesis or alternate hypothesis is accepted. The rationale behind using a non-parametric test is because the distribution of evidence from both systems does not follow a normal distribution (Shapiro-Wilk test on emotion evidences: pValue < 0.05 for all the 6 emotions in both systems). The results from this experiment are tabulated below:

Table 4.2: Results of Kruskal-Wallis rank sum test between different data populations in Emotient and Affectiva (INTRASYSTEM)

Profession	Emotion	Emotient		Affectiva	
		p-Value	Decision	p-Value	Decision
CEO vs REST	joy	0	Rejected	0	Rejected
	anger	0	Rejected	0	Rejected
	surprise	0	Rejected	0	Rejected
	fear	0	Rejected	0	Rejected
	disgust	0	Rejected	0	Rejected
	sadness	0	Rejected	0	Rejected
Politician vs REST	joy	0	Rejected	0	Rejected
	anger	0	Rejected	0	Rejected
	surprise	0	Rejected	0.6203	Not Rejected
	fear	0	Rejected	0	Rejected
	disgust	0	Rejected	0	Rejected
	sadness	0	Rejected	0	Rejected
Spokesperson vs REST	joy	0	Rejected	0	Rejected
	anger	0	Rejected	0	Rejected
	surprise	0	Rejected	0	Rejected
	fear	0	Rejected	0	Rejected
	disgust	0.7008	Not Rejected	0	Rejected
	sadness	0	Rejected	0.6903	Not Rejected

We can observe from 4.2 that the data distribution is significantly different in the comparison of CEO vs REST in both systems and the null hypothesis is rejected for all the emotions. While comparing the distribution of spokesperson vs rest, the data distribution is not significantly different for disgust in Emotient and sadness in Affectiva. The null hypothesis is accepted for both these cases. Similarly, for the comparison of politicians vs rest, the null hypothesis is not rejected for surprise in Affectiva. Further, the same analysis is extended to inter-system comparison, wherein the emotion distribution of a subject with a particular profession is compared between Emotient and Affectiva. For instance, the emotion evidence distribution of CEOs from Emotient is compared with the emotion evidence distribution of CEOs in Affectiva. The hypothesis for the analysis is presented below:

H_0***: The distribution of*** e_i ***for each*** p_i ***is same between*** AER_i
H_A***: The distribution of*** e_i ***for each*** p_i ***is not same between*** AER_i,where
$e_i \in joy, anger, surprise, fear, disgust, sadness,\ AER_i \in Emotient, Affectiva,$
$p_i \in CEO, Politician, Spokesperson$

Table 4.3: Results of Kruskal-Wallis rank sum test between different data populations between Emotient and Affectiva (INTERSYSTEM)

Profession	Emotion	pValue	Decision
CEO	joy	0	Rejected
	anger	0	Rejected
	surprise	0.7047	Not Rejected
	fear	0	Rejected
	disgust	0	Rejected
	sadness	0	Rejected
Politician	joy	0	Rejected
	anger	0	Rejected
	surprise	0	Rejected
	fear	0	Rejected
	disgust	0	Rejected
	sadness	0	Rejected
Spokesperson	joy	0	Rejected
	anger	0	Rejected
	surprise	0	Rejected
	fear	0	Rejected
	disgust	0	Rejected
	sadness	0	Rejected

From the inter-system comparisons, it can be noted that the null hypothesis is accepted only in the case of Surprise for CEOs showcasing the significant difference in the data distribution across professions. Results from intra-system and inter-system comparisons indicate the statistical difference between the automatic emotion recognition systems based on the individual's profession and are in agreement with the study presented by (47).

4.2 Head Orientation

From the curated dataset, the systems produce the output for each of the frames. For every frame, evidence for each emotion and the head orientation is estimated synchronously. Both systems involve a different pipeline to estimate the head orientation in terms of Euler angles. The following analysis compares the estimated Euler angles from both systems and investigates the second research question: "Is the computation of head orientation independent of the facial recognition systems used?".

4.2.1 Comparison of Estimated Euler Angles

The overall output from both systems is aggregated and there are 1099009 frames processed by both systems. The Euler angles are visualised as a time series (figure 4.4 and the agreement of Euler angles estimated from both systems are visualised as a scatter plot presented in the figure 4.5.

Figure 4.4: Distribution of Euler Angles in both systems near common agreement index

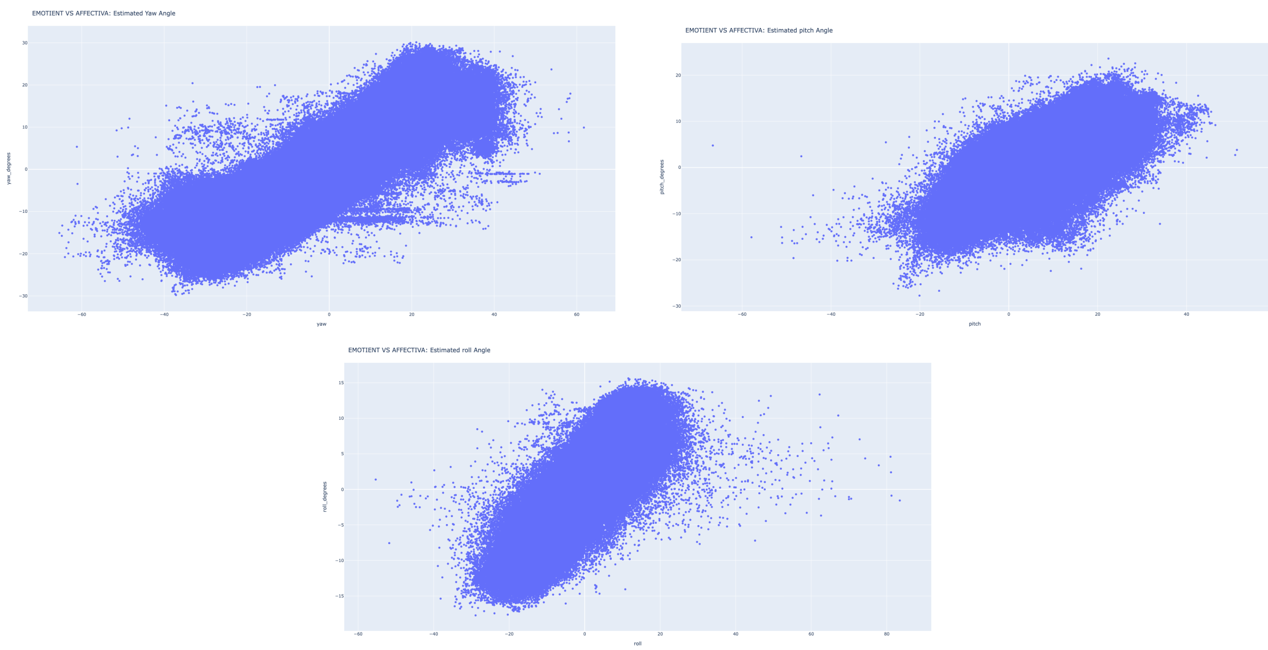

Figure 4.5: Euler Angles from Affectiva vs Euler Angles from Emotient

Table 4.4: Descriptive Statistics of Euler angles from both systems

Emotient (nobs = 1099009)						
	mean	**std**	**range**	**variance**	**skewness**	**kurtosis**
yaw_degrees	0.240	9.660	59.940	93.390	0.100	-0.500
roll_degrees	0.500	4.970	33.350	24.710	-0.240	-0.090
pitch_degrees	1.190	5.500	51.330	30.300	-0.250	-0.080
Affectiva (nobs = 1099009)						
	mean	**std**	**range**	**variance**	**skewness**	**kurtosis**
yaw	-0.860	14.400	127.150	207.460	-0.020	0.160
roll	-0.180	7.460	138.790	55.590	-0.050	0.770
pitch	5.920	8.850	117.930	78.300	0.200	-0.020

The descriptive statistics of the estimated Euler angles from both systems are compared in the table 4.4. In most parts of the videos, the subjects are looking directly at the camera. The subject while delivering a speech may be using a teleprompter or the video footage may be captured using more than one camera. Thus, when , there may be a difference in the camera perspective which may reflect in the estimated Euler angles. The yaw degrees which capture the head orientation in the horizontal axis have a mean of 0.24 from Emotient while the mean yaw angle from Affectiva is -0.86. We can observe that there is a variation of the mean for each of the Euler angles between the systems. T-Test is used to statistically compare the mean Euler angles from both systems and the results are tabulated in 4.5.

Comparison of Mean Euler Angles

T-test comparison of mean Euler angles is done at two levels: intra-system and inter-system. Initially, the overall mean for each Euler angle is compared between the systems, the mean is compared with reference to the profession, i.e, the mean of CEOs from Emotient is compared with the mean of CEOs from Affectiva. This analysis is then extended to the intra-system where the mean angle for CEOs in Emotient is compared with the mean angles of Politicians and Spokespersons (NOT CEO). The results of this analysis are tabulated in the tables 4.5 and 4.6.

From table 4.5 and 4.6 we can observe a significant difference in the mean for all the comparisons ($pValue < 0.001$) . A similar analysis is extended for gender, and a significant difference in the mean is observed for both intra-system and inter-system. (The results are tabulated in table 1 and included in appendix). Both systems estimate the Euler angles from the same set of videos. From this analysis, we are able to observe that for the mean of estimated Euler angles is significantly different across the profession and systems.

Comparison of Distribution of Euler Angles

The distribution of Euler angles for the overall aggregated dataset is compared using histograms. To generate the histograms, the Euler angles are normalized between 0 to 1. The Euler angles from both systems are estimated in terms of radians and each Euler angle has a different range.

Table 4.5: Results of T-test between different data populations, Emotient vs Affectiva (INTER-SYSTEM)

Profession	Euler Angle	T-statistic	pValue
Overall	yaw	66.4722	0
	pitch	-475.0066	0
	roll	79.7411	0
CEO	yaw	-45.6506	0
	pitch	-169.476	0
	roll	-14.2951	2.46E-46
	yaw	43.43487	0
Politician	pitch	-395.121	0
	roll	66.36111	0
	yaw	94.02979	0
Spokesperson	pitch	-210.064	0
	roll	65.86061	0

Table 4.6: Results of T-test between different data populations, Within Emotient and Affectiva (INTRA-SYSTEM)

Profession	**Euler Angle**	**T-statistic**	**pValue**	**T-statistic**	**pValue**
	yaw_degrees	112.8348	0	158.5085	0
CEO vs REST	pitch_degrees	-4.0756	4.59E-05	-25.4543	6.96E-143
	roll_degrees	-61.3171	0	10.6930	1.10E-26
	yaw_degrees	-8.5659	1.07E-17	20.3849	2.37E-92
Politician vs REST	pitch_degrees	-51.7213	0	-26.8725	5.19E-159
	roll_degrees	62.6402	0	38.9556	0
	yaw_degrees	-75.2017	0	-142.9412	0
Spokesperson vs REST	pitch_degrees	62.0236	0	49.8793	0
	roll_degrees	-24.9734	1.30E-137	-52.4821	0

Once normalized, the Euler angles are segregated into specific bins facilitating comparison between histograms. Thus, all the presented histograms share the x-axis(normalized Euler angles) and the relative percentage of the number of frames in each bin is plotted on the y-axis.

Figure 4.6 indicates the distribution of Euler angles for both systems. From the distribution of Euler angles from both systems, it seems that the histogram from Affectiva is fairly symmetrical and the histograms from Emotient are moderately skewed. It appears that Emotient is sensitive to a small change in head positions and able to discriminate different head positions across frames more effectively than Affectiva. But these histogram plots provide little information on the agreement between the two systems. Both the systems detect the head orientation frame-by-frame and this estimation is not dependent on previous frames [(6)]. Since the same video is given as input to both systems, the results are expected to be comparable. However, the head pose estimation is dependent on other components of the system architecture where the algorithms employed to estimate facial features(face bounding box and facial landmarks) are different. This difference in the mean of Euler angles is acknowledged and further analysis is done to understand whether the Euler angles have the same distribution and to inspect the agreement between the systems. The hypotheses for this analysis are as follows:

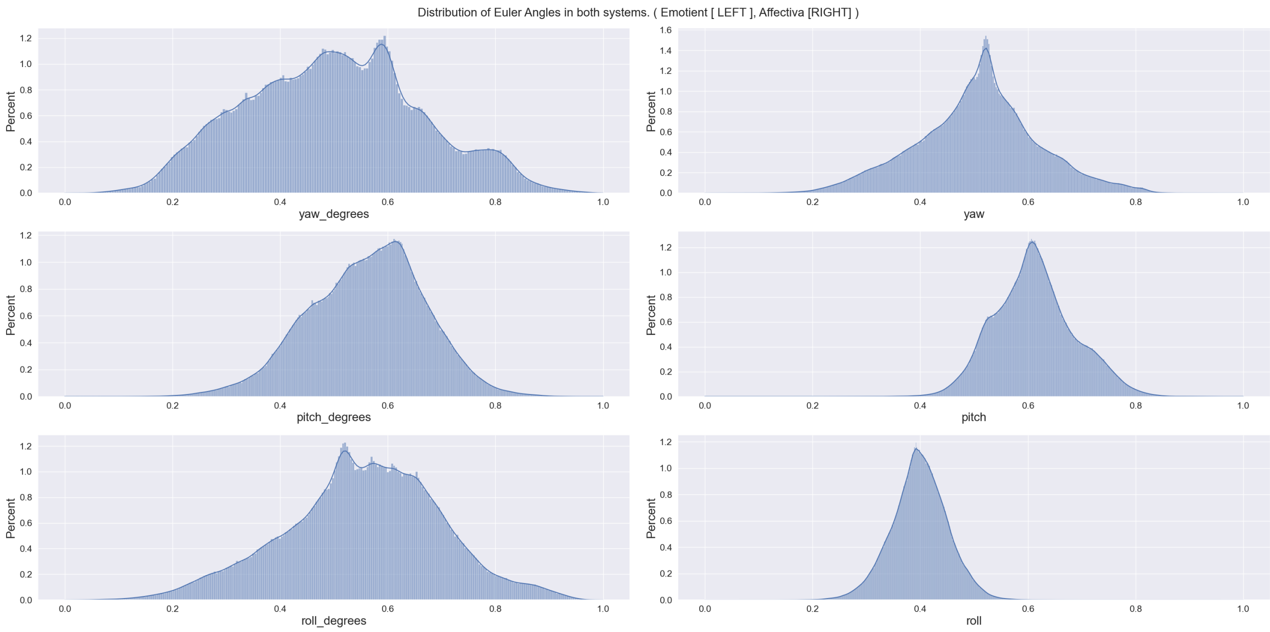

Figure 4.6: Distribution of Euler Angles in both systems, Emotient(LEFT), Affectiva(RIGHT)

H_0***: The distribution of*** $EulerAngle_i$ ***is same for*** AER_i
H_A***: The distribution of*** $EulerAngle_i$ ***is not same for*** AER_i,where
$EulerAngle \in Yaw, Pitch, Roll$, $AER \in Emotient, Affectiva$

Table 4.7: Comparison of the distribution of Euler angles between Emotient and Affectiva

	yaw		pitch		roll	
	spearman	kruskal	spearman	kruskal	spearman	kruskal
coefficient	0.882	3439.2291	0.6977	179685.8	0.832	6631
pValue	0	0	0	0	0	0

From the table above, we can observe that there is a strong correlation between the Euler angles. The yaw and roll angles have a strong correlation and the pitch angle which captures the position of the head on the vertical axis is moderately correlated. All the correlations are statistically significant ($pValue < 0.01$). However, we can see that the distribution of Euler angles between the systems is significantly different(Kruskal-Wallis Test: $pValue < 0.01$). Thus, the null hypothesis is rejected. The analysis is further extended to include the profession dimension and the results are summarised in table 4.8.

From the table above, we can observe the variation in the strength of correlation for pitch between systems with respect to the profession. A strong correlation of Pitch angle is observed for a spokesperson(0.78), a moderate correlation for politicians(0.68), and the lowest correlation for CEOs(0.53). However, the difference in the distribution is still significant irrespective of the profession.

Table 4.8: Comparison of the distribution of Euler angles between Emotient and Affectiva with reference to Profession

Occupation	Euler Angle	spearman	pValue	kruskal	pValue
	yaw	0.88792963	0	4753.34396	0
CEO	pitch	0.53316943	0	22756.0773	0
	roll	0.86333495	0	99.0934704	2.41E-23
	yaw	0.87625127	0	2035.19066	0
Politician	pitch	0.68619733	0	125046.571	0
	roll	0.83258514	0	5413.81508	0
	yaw	0.89381538	0	7013.2009	0
Spokesperson	pitch	0.77916791	0	34269.5775	0
	roll	0.81701773	0	3011.62577	0

Variation with emotions

Table 4.9 indicates the correlation between the distribution of Euler angles during episodes of a specific emotion from both the systems. We can observe that there is a strong correlation between all the emotions in the case of yaw and roll. There is a variation in the strength of correlation for pitch angle, wherein for emotions joy and anger we are able to see a moderate correlation. If the difference in the distribution is also statistically significant for all the emotions ($pValue < 0.001$) except for the comparison of yaw for the emotion surprise and comparison of roll for the emotion sadness.

Table 4.9: Comparison of the distribution of Euler angles between Emotient and Affectiva with reference to emotion. $pValue < 0.01$ for all comparisons except observations highlighted in bold with *.

	yaw		pitch		roll	
	Spearman	kruskal	Spearman	kruskal	Spearman	kruskal
joy	0.77	31.84	0.58	12701.19	0.84	160.78
anger	0.87	386.07	0.65	11541.27	0.81	597.56
surprise	0.90	**0.66***	0.71	2697.21	0.80	85.85
fear	0.84	16.63	0.76	527.66	0.86	32.86
disgust	0.87	3200.27	0.79	22683.72	0.84	1276.15
sadness	0.86	91.28	0.76	765.15	0.86	**0.64***

From the inter-system comparisons, we can observe a significant difference in the mean Euler angles, their distribution, and the estimated head orientation varied with respect to emotions. Interestingly, the difference in the mean and the data distribution can be observed within systems and are consistent across professions. These results indicate the estimation of head orientation is dependent on the facial recognition systems. The following analysis investigates the relationship of head orientation with emotions.

4.2.2 Head Direction and Emotions

The following analysis presents the idea of visualising the change in the emotion intensity and the corresponding change in head orientation as a time series. The video of a prominent politician during a coronavirus briefing is selected for analysis and then generalised over the entire dataset. The output from both the systems after processing contains 21653 frames, out of which Emotient predicts the subject was angry in 30% of the frames and Affectiva predicts the subject was angry in 50% of the frames. However, the evidence values for anger vary frame-by-frame. From the figure 4.7, it is evident that there is a variation in anger scores between the systems. The red reference lines indicate the media time where the variation in anger was greater than 90%. These media times are the indicator indices where one system recognizes anger with the highest confidence while the other system recognizes little evidence of anger.

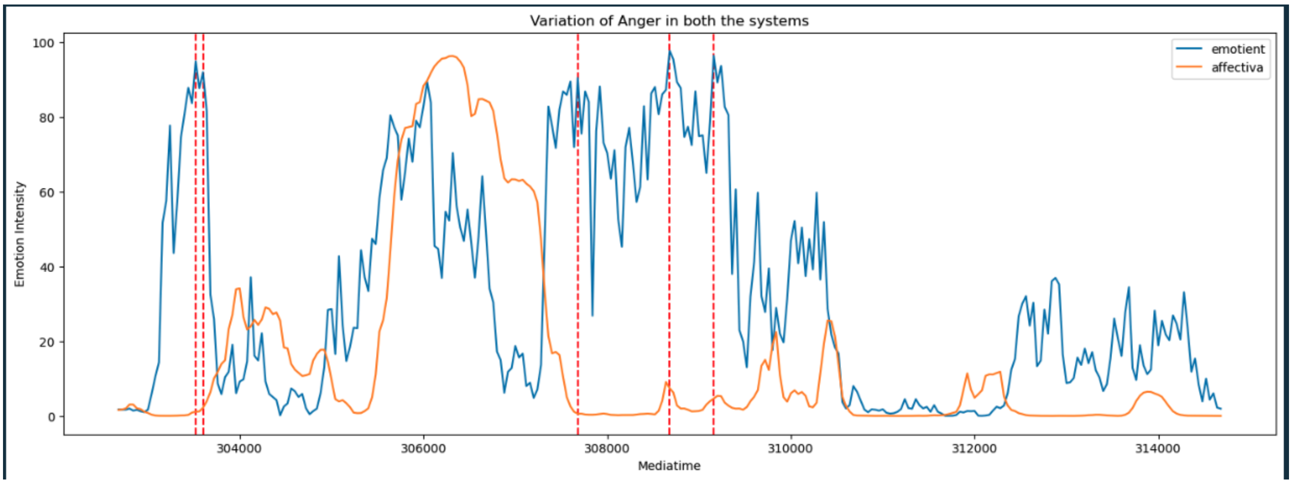

Figure 4.7: Variation of Anger in both the Systems

The rationale behind this analysis is to compare the head pose estimated by the systems where there is significant evidence of emotion. As highlighted in the figure 4.7, there is variation in anger score for each frame. Thus, a common agreement index is to be evaluated where the anger score from both systems is greater than the activation threshold. The activation threshold for this analysis is considered 90%. The variation of anger intensities at two such common agreement indices is presented in the figure 4.8. The dataset is filtered for 300 frames on either side of the common agreement index and the variation of Euler angles from both the systems is visualized.(Figure 4.4). This time series analysis is appropriate for a single subject. Considering the change in emotional state or the change in head orientation for the entire dataset will not be appropriate as the data from multiple subjects displaying different emotions are aggregated. Thus, we need a technique to capture the trend in the overall aggregated dataset. Both systems estimate the head orientation in terms of Euler angles. Yaw degree captures the head position in the horizontal axis. A negative value of the yaw degree indicates that the head is positioned to the left and the positive value of the yaw angle indicates the head is positioned to the right. Similarly, a positive pitch indicates whether the head is pointing up or whether the head is pointing downwards. The roll angle captures the head tilt; positive

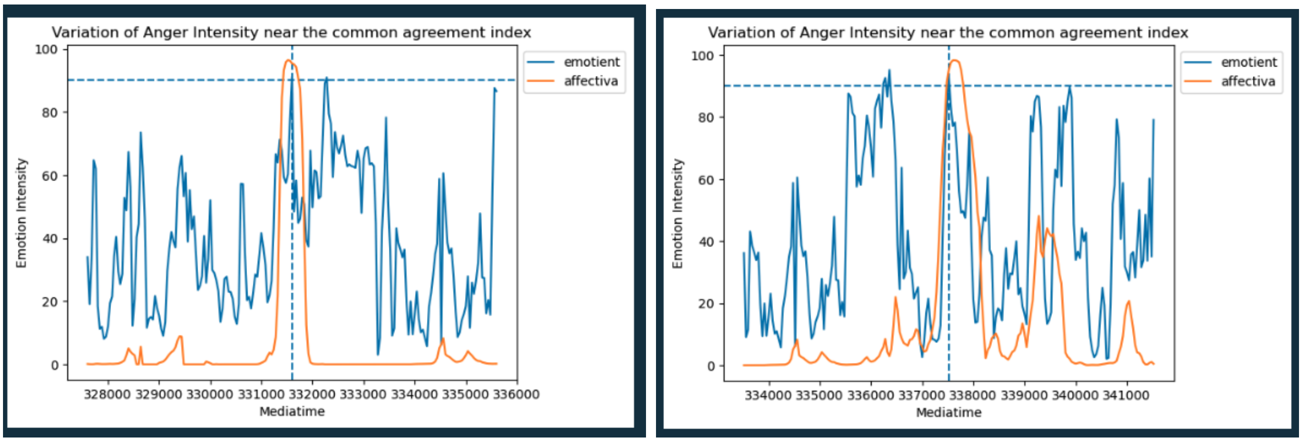

Figure 4.8: Variation of Anger Intensity at Common agreement Index

values for left tilt and negative values for right tilt. Thus, the head direction on both the vertical and horizontal axis can be derived from the sign of the Euler angles. The magnitude of the Euler angles indicates the deviation of the head on the axis of interest from the reference.

The following analysis compares the contingency table to understand the agreement between the systems in estimating the head direction when the subject is expressing a specific emotion. Initially, the contingency table is created for the entire dataset without considering any emotions. The rationale behind this analysis is to understand whether there is any disagreement in the estimated head direction for the given frame by the systems and does this differs based on the underlying emotion. Figure 4.9 presents the contingency table for the three Euler angles from both systems. We can observe that there is a high agreement between the systems for all three angles. In the case of yaw which captures the head orientation in the horizontal axis, the systems have an agreement of 86%. However, for 8% of frames, Emotient estimated the yaw angle to be positive (head is to the right) while Affectiva estimated a negative angle(head to the left). Conversely, for 6% of the frames, Affectiva estimated a positive yaw angle while Emotient gave a negative yaw angle. Similarly, for the Pitch angle, we can see a good agreement of 76% between the systems. For 19% of the frames, we can observe contradicting estimations from the systems; Emotient estimated that the head is pointing up from the reference line, and Affectiva suggests that the head is pointing down. The estimations of roll angle between the systems have a high agreement of 82%. These Euler angles are predicted through a regression pipeline and the architecture differs in both these systems. The residual error which is the intrinsic part of the regression pipeline is cascaded to each of these estimations. But, this presents an interesting question to explore the impact of emotions in this agreement between the systems. Ideally, the agreement between the systems on the head direction of the subject on different axes should be independent of the underlying emotion. Figure 4.10 and 4.11 indicate the contingency table between systems when the subjects are showing evidence of joy and anger. From the figure 4.10, we can observe that when our subjects are showing evidence of joy, there is an 84% agreement in yaw degrees, 79% in pitch, and 83% in the roll. However, when subjects are showing evidence of anger, the disagreement between the systems increases significantly. We can observe that the systems agree only for 59% of the frames in the vertical axis (pitch).

Statistical significance of the agreement between the systems

To understand the statistical significance of the agreement between the systems, cohen-kappa scores are calculated between the systems for the estimated head direction. Cohen-kappa score expresses the level of agreement between the systems. Cohen-kappa statistic is calculated using the below-mentioned formula, where p_o is the observed agreement ratio and p_e is the expected agreement ratio.

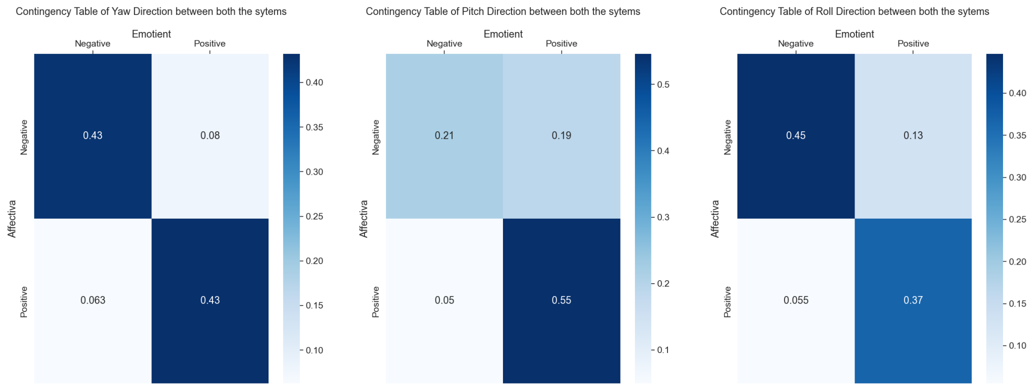

Figure 4.9: Contingency table of estimated Euler angles between Emotient and Affectiva

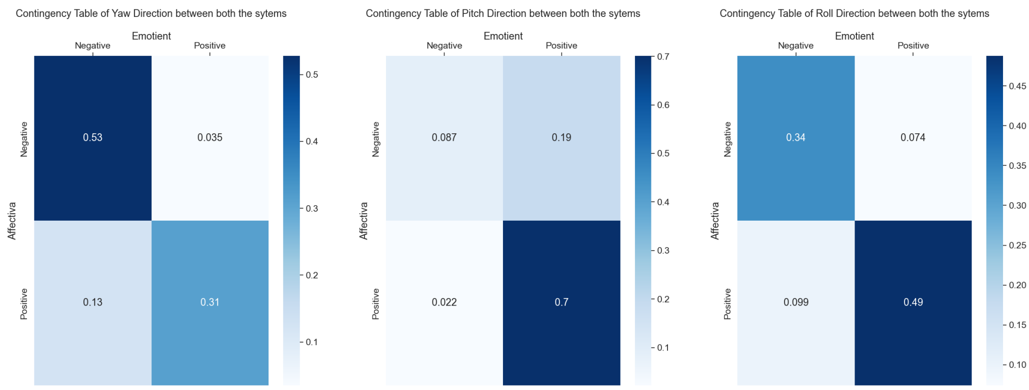

Figure 4.10: Contingency table of estimated Euler angles between Emotient and Affectiva for Joy

Figure 4.12 is a plot that highlights the variation of head direction agreement between Emotient and Affectiva for different emotions. We can observe that the agreement between the systems for pitch degrees is lowest for anger and highest for fear. Similarly, the agreement during joy is highest for surprise and lowest for fear. For anger, we can observe that there is a lower agreement for all the Euler angles than the overall agreement and for the emotion of surprise, we have a higher agreement for all the Euler angles than the overall agreement. Additionally, all the values are greater than 0 indicating that the agreement between the systems is not by chance.

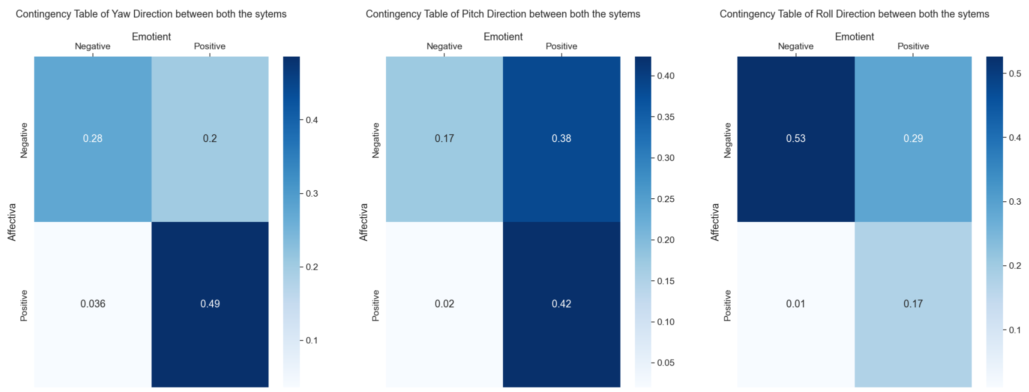

Figure 4.11: Contingency table of estimated Euler angles between Emotient and Affectiva for anger

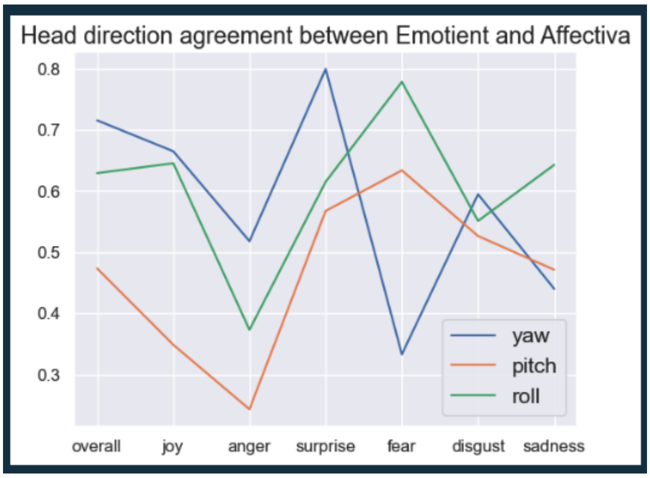

Figure 4.12: Variation of Cohen-Kappa agreement between the systems for different Emotions

4.2.3 Head Orientation and Emotions

By just considering the direction of the head, a lot of information about the head orientation is abstracted. When the head direction takes a binary value in each axis, there is no discrimination between the low and high values of Euler angles. For instance, the head direction is 'right' when the yaw degree is positive irrespective of whether the magnitude is +0.5 or +20. The following analysis classifies the Euler angles into three predefined ranges, and explores the interaction of head direction and perception of different emotions. Both systems represent the head orientation in terms of Euler angles, and we can observe that the range of each Euler angle is different. The series of steps followed to perform this analysis is detailed below.

As part of the research work, a python script is developed that takes in the aggregated data from both the systems as the input and creates a bubble chart that compares the head orientation between systems during episodes of different emotions. The script implements the above-mentioned algorithm. As the initial step, the distribution of the Euler angles is considered to determine the range of angles which is then used for evaluation. In the curated dataset, the outliers are removed in the post-processing phase. The yaw degree which captures the head

Algorithm 1 To analyse the interaction of head direction and perception of different emotions

for $angle = yaw, pitch, roll$ **do**
    Find the max and min angle from its distribution such that it covers 95% of the datapoints
    The range is computed as [minvalue,-5),[-5,5),[5,maxvalue]
end for
yawranges = getYawRange()
pitchranges = getPitchRange()
for $yawrange = [...yawranges]$ **do**
    for $pitchrange = [...pitchranges]$ **do**
        Filter the frames where the estimated Euler angle is in the yaw-range and pitch-range
        Evaluate the relative frequency of such frames
    end for
end for

position in the horizontal axis has a range of (-30 to +32) in Emotient and (-34 to +35) in Affectiva. By considering the histogram of the yaw degrees from both systems, 96% of the data is encapsulated in the range of [-26 to 30]. Similarly, the pitch angle which captures the head position in the vertical axis has a range of (-29 to 32) in Emotient and (-32 to +35) in Affectiva. A major chunk of frames(97%) are in the range of [-22.5 to +16.5]. Thus, these values are used to segregate the angles into three categories for further analysis.

The data frames are filtered based on the emotion evidence from each system. This step is done to ensure that only the frames with concrete evidence of a specific emotion are selected. The activation threshold for the analysis is set to be 90%. For instance, when filtering the frames based on anger, only the frames with anger intensity greater than 0.9 from Emotient and the frames with anger score greater than 90% in Affectiva are filtered in. The relative frequency with reference to the emotion is tabulated in Figure 4.13 and Figure 4.14.

	yaw_range	pitch_range	#frames	anger	joy	fear	disgust	sadness	surprise
0	-26 to -5	-22 to -5	31006	225 (0.7%)	778 (2.5%)	806 (2.6%)	2744 (8.8%)	910 (2.9%)	1753 (5.7%)
1	-26 to -5	-5 to 5	175329	698 (0.4%)	15583 (8.9%)	5673 (3.2%)	5495 (3.1%)	2018 (1.2%)	8454 (4.8%)
2	-26 to -5	5 to 16.5	95659	748 (0.8%)	11014 (11.5%)	2176 (2.3%)	2905 (3.0%)	594 (0.6%)	2095 (2.2%)
3	-5 to 5	-22 to -5	55059	2465 (4.5%)	1344 (2.4%)	1764 (3.2%)	9776 (17.8%)	266 (0.5%)	2220 (4.0%)
4	-5 to 5	-5 to 5	202511	2327 (1.1%)	22781 (11.2%)	16541 (8.2%)	11541 (5.7%)	3846 (1.9%)	7184 (3.5%)
5	-5 to 5	5 to 16.5	95116	667 (0.7%)	11186 (11.8%)	2424 (2.5%)	2801 (2.9%)	850 (0.9%)	1589 (1.7%)
6	5 to 30	-22 to -5	53162	659 (1.2%)	2573 (4.8%)	1177 (2.2%)	4464 (8.4%)	218 (0.4%)	1797 (3.4%)
7	5 to 30	-5 to 5	210138	2307 (1.1%)	18173 (8.6%)	12419 (5.9%)	14345 (6.8%)	933 (0.4%)	5875 (2.8%)
8	5 to 30	5 to 16.5	38710	587 (1.5%)	3584 (9.3%)	791 (2.0%)	5344 (13.8%)	249 (0.6%)	544 (1.4%)

Figure 4.13: Distribution of frames across emotions for a specific range of Euler angles in Emotient

Each row in the table indicates the distribution of frames across emotions in the particular range of Euler angles. For instance, the results in the first row of Table 4.13 depict that, 31006 frames in the aggregated dataset have yaw degrees in the range of -26 to -5 and pitch degrees in the range of -22 to -5. This indicates that the head is oriented in the left-down direction in the 2-dimensional space. Further, we can observe the relative distribution of frames with reference

	yaw_range	pitch_range	#frames	anger	joy	fear	disgust	sadness	surprise
0	-26 to -5	-22 to -5	37703	633 (1.7%)	86 (0.2%)	0 (0.0%)	442 (1.2%)	382 (1.0%)	1572 (4.2%)
1	-26 to -5	-5 to 5	95248	1290 (1.4%)	804 (0.8%)	13 (0.0%)	1057 (1.1%)	1243 (1.3%)	3454 (3.6%)
2	-26 to -5	5 to 16.5	110205	2686 (2.4%)	2100 (1.9%)	0 (0.0%)	549 (0.5%)	546 (0.5%)	926 (0.8%)
3	-5 to 5	-22 to -5	56885	568 (1.0%)	540 (0.9%)	46 (0.1%)	803 (1.4%)	630 (1.1%)	1602 (2.8%)
4	-5 to 5	-5 to 5	127886	5245 (4.1%)	2224 (1.7%)	200 (0.2%)	1611 (1.3%)	901 (0.7%)	5585 (4.4%)
5	-5 to 5	5 to 16.5	96538	3546 (3.7%)	4069 (4.2%)	27 (0.0%)	712 (0.7%)	766 (0.8%)	1365 (1.4%)
6	5 to 30	-22 to -5	9494	65 (0.7%)	158 (1.7%)	0 (0.0%)	282 (3.0%)	104 (1.1%)	617 (6.5%)
7	5 to 30	-5 to 5	109249	3388 (3.1%)	2048 (1.9%)	53 (0.0%)	2043 (1.9%)	677 (0.6%)	4288 (3.9%)
8	5 to 30	5 to 16.5	136509	2885 (2.1%)	3751 (2.7%)	13 (0.0%)	1187 (0.9%)	726 (0.5%)	1395 (1.0%)

Figure 4.14: Distribution of frames across emotions for specific range of Euler angles in Affectiva

to each emotion. Out of the 31006 frames, we only have 225 frames corresponding to anger which accounts for 0.7%. Emotions like surprise and disgust have the highest frequency of such frames in the specified orientation of the head. In comparison, for the same range of Euler angles, Affectiva produces 37703 frames 4.14 which is 20% more than Emotient. However, we can see a difference in the relative frequency of different emotions. Except for anger, all other emotions have a lower frequency of frames in the range of Euler angles, and no frames for fear were found with concrete evidence in these filtered frames. 3.3

The results presented in the above tables are visualized as a bubble chart. The yaw degrees and the pitch degrees are classified into three categories. For the purpose of visualization, these three categories are encoded into discrete values. Figure 4.15 and 4.16 presents the comparison of bubble charts for different emotions. There is a clear indication of a specific head orientation during episodes of different emotions.

The charts from Emotient for anger indicate that a huge chunk of frames is distributed in the negative direction of pitch and movement of the head along the horizontal axis is minimal. The chart for joy indicates that the head is predominantly in the positive direction of pitch and is dispersed over the horizontal axis. We can see that the result for joy from Affectiva matches with that of Emotient, where there is a significant distribution of frames in the positive axis of pitch, and the frequency is dispersed along the horizontal axis. However, we can observe a disagreement between the systems for anger. A major chunk of frames is in the positive axis of pitch contradicting the results from Emotient. A comparison of head orientation along both axis for all the emotions is presented in the figure 4.15 and 4.16.

From the presented analysis, we can observe that there is evident agreement on the head direction estimated by the systems. To explore the agreement between the systems, the variation of Euler angles and the interaction between them are analyzed. 4.15 and 4.16 illustrate that both Emotient and Affectiva agree on the association of positive pitch angle with emotion joy, however, there is an evident variation in agreement for other emotions. The results from bubble charts provide a foundation to explore the association of specific head orientation with each emotion.

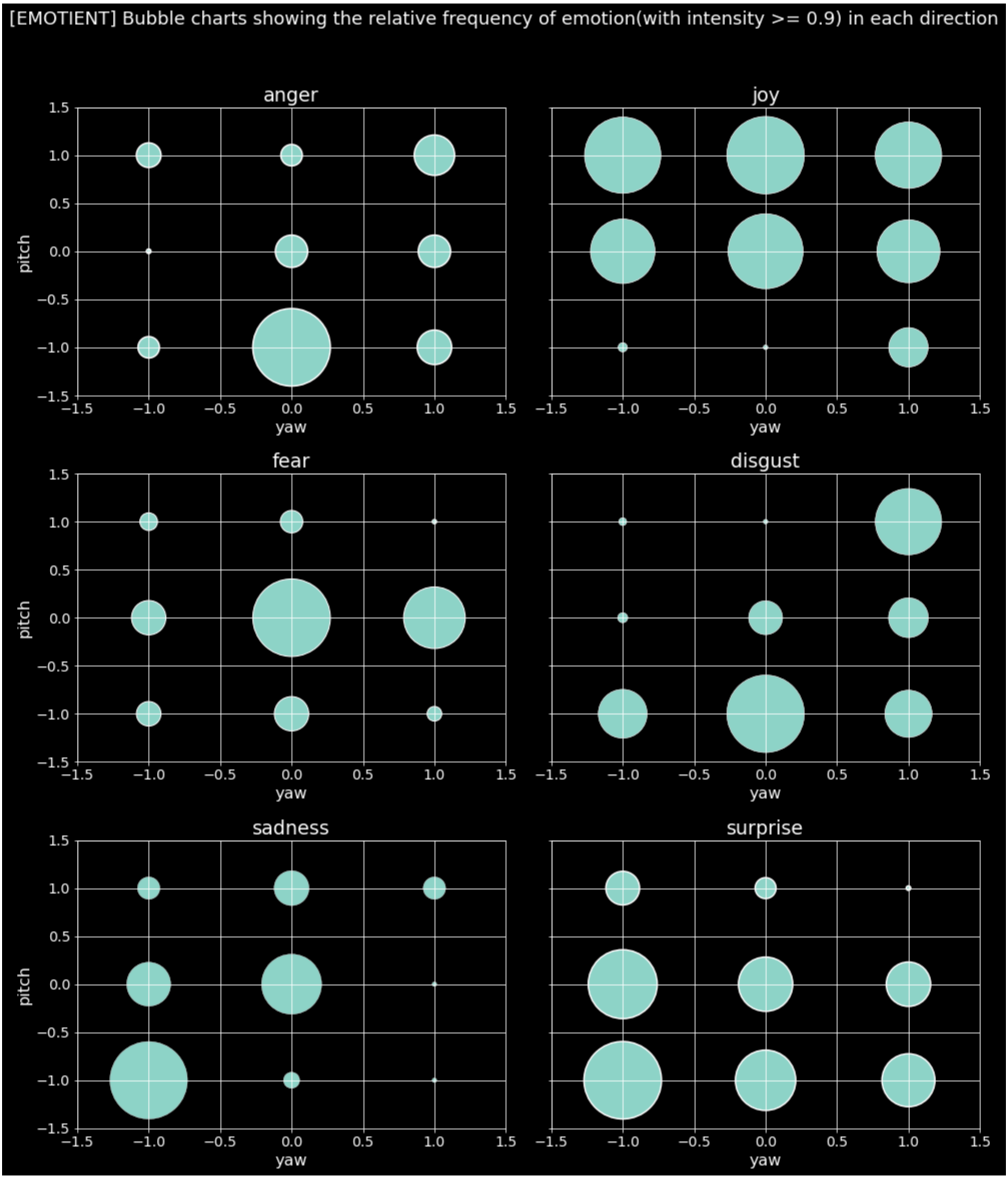

Figure 4.15: Variation of head orientation in Emotient with respect to Emotions

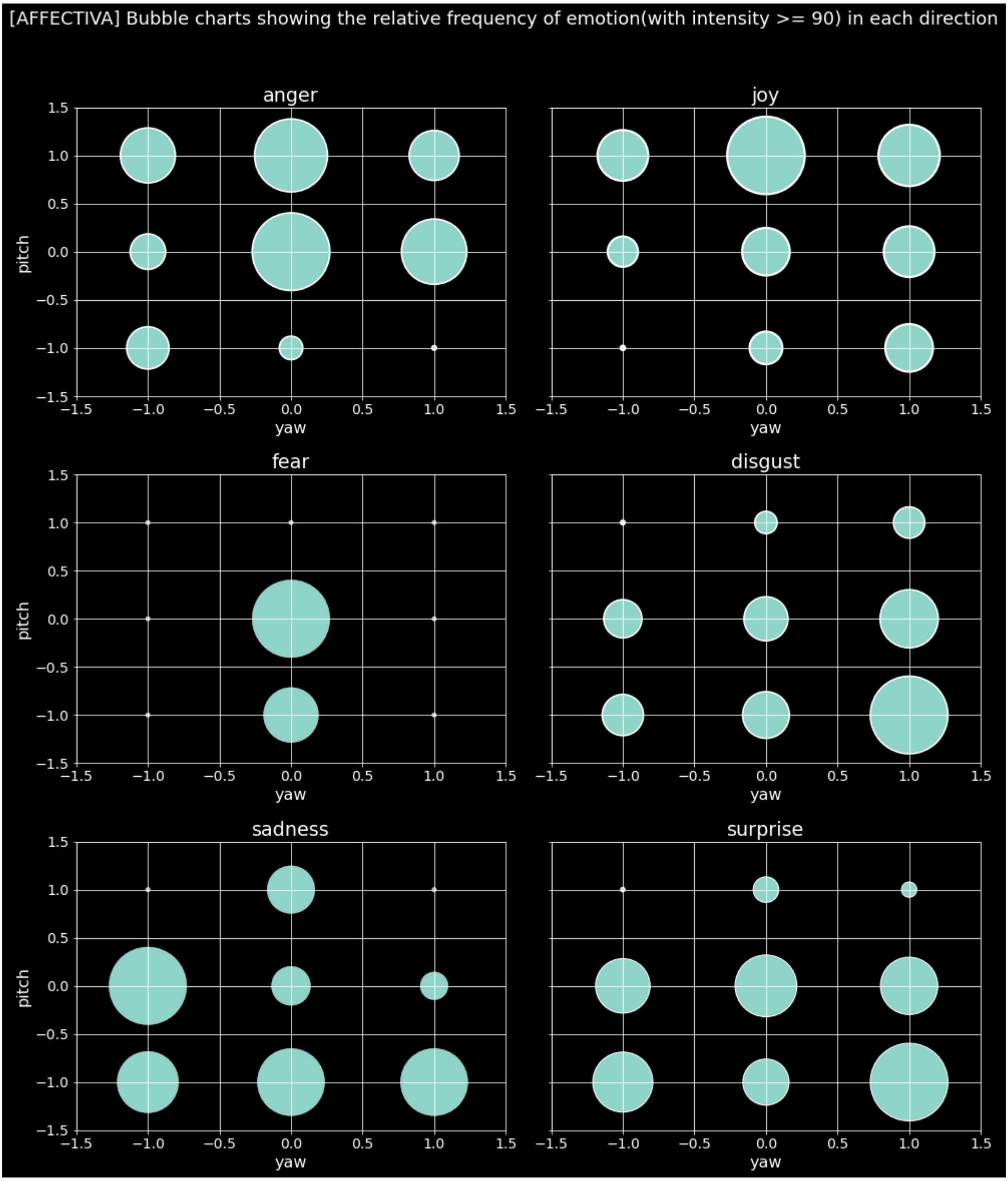

Figure 4.16: Variation of head orientation in Emotient with respect to Emotions

4.2.4 Regression Analysis

The following section describes the regression analysis. The initial analysis investigates the possibility of deriving head orientation given the emotional state. For instance, will the subject look down when he is angry? To explore this aspect, the Euler angles (head orientation) are considered as a function of emotion intensities, and the relationship is evaluated by estimating the R-squared value. R-squared value gives us an estimate of how much variance in the data can be explained by the independent variables. Further, the analysis is extended to understand if the emotion state can be represented as a function of Euler angles.

Head Orientation as a Function of Euler angles This section explores the premise that the head orientation can be represented as a function of the Euler angle. The following steps are performed as pre-processing steps for regression analysis. The frames where there is a clear indication of a particular emotion are identified by filtering our frames with an intensity less than the activation threshold. The activation threshold for this analysis is set at 90%. All the Euler angles are normalized using mix max scaling. Then, the OLS regression analysis is performed for anger and joy against each Euler angle. The emotions are considered as independent variables and the Euler angles are considered as dependent variables. The relationship between Euler angles and emotional intensity is evaluated using the R^2 (uncentered) score and the results are tabulated below.

Table 4.10: Comparison of R^2 Score estimates (uncentered) for linear models capturing the relationship between Euler angles(Head Orientation) and Emotion Intensity

System	Emotion	Yaw	Pitch	Roll
Emotient	anger_intensity	0.544	0.533	0.596
	R^2	0.919	0.89	0.95
	joy_intensity	0.498	0.5904	0.5441
	R^2	0.895	0.962	0.923
Affectiva	anger	0.51	0.58	0.42
	R^2	0.878	0.941	0.933
	joy	0.54	0.49	0.38
	R^2	0.89	0.918	0.95

The emotion values are able to explain a lot of variance in the data and the coefficient values for all linear models are statistically significant ($pValue < 0.001$). The complete OLS regression results are attached in the appendix section. From the table above, we can observe that there is a clear indication of a relationship between head orientation and emotions.

Emotions as a function of Euler angles The rationale behind this analysis is to understand if head orientations need to be included in the AU intensity pipeline. (7) and (24) illustrate the pipeline of the considered automatic emotion recognition systems and the head orientation is not considered during the estimation of emotion evidence in each frame. However, the systems depend on facial landmarks to detect emotional intensity. When the head is at an extreme angle, there is a possibility for obscuration of facial landmarks. (67). This may lead to intensification or deintensification of emotional intensity. This idea can be explored in the future aspects of the research project. (8) in their latest developments have included the head orientation in their emotion evidence detection pipeline which provides the required initial validation to explore it in detail.

4.2.5 Fuzzification of Euler Angles

The following analysis presents the idea of fuzzification of Euler angles to understand the relationship between head orientation on the perception of emotions by AERs. The analysis presented so far considered the Euler angles as a discrete variable and each Euler angle has been divided into three categories based on its sign and magnitude. We can use the sign of the Euler angles to determine the head direction and in this section, the head direction is further divided into subcategories. For instance, if the value of the yaw degree is positive, it indicates that the head is to the right. However, we need to discriminate between head direction when the yaw degrees are +0.5 and +20. Thus, the head direction in the horizontal axis is further divided into the following categories: Extreme-left, significant-left, left, neutral, right, significant-right, and extreme-right. Similarly, this pattern of bifurcation of head direction is applied to the vertical axis (pitch degrees) and the head tilt (roll degrees).

The head direction is not determined by simply taking different ranges of Euler angles as in the previous analysis, since it may not be appropriate because a crisp limit will be introduced into the system wherein yaw degrees of -4.99999 may be considered as left and -5 will be considered as significant-left. Thus, to effectively determine the head direction, a fuzzy membership function is created that takes in the Euler angles, eliminates the crisp limits, and produces the head direction in each of the axes. The steps followed in the fuzzification of Euler angles are detailed below.

Algorithm 2 Fuzzification of Euler angles

```
for angle = yaw, pitch, roll do
    angleRange = getRange(distribution)
    angle-fuzzy-membership-values = trimf(angleRange)
    for emotion in targetEmotions do
        computeCohenKappaScore()
        computeRelativeFrequency()
        plotBubbleChart()
    end for
end for
```

The core idea of this analysis is the fuzzification of the Euler angles. The fuzzy membership function is used to convert the crisp input of Euler angles to a fuzzy set. The rationale behind this analysis is to create a common basis of head direction that facilitates comparison between the systems. The fuzzy membership function maps each Euler angle in the dataset to a value between 0 and 1, which is the degree of membership. This value quantifies the grade of membership to the target fuzzy set. Unlike the previous analysis, which only considered the sign of the Euler angles to determine the head direction, this fuzzy membership function will also create a sense of the magnitude of the Euler angles. For instance, the measurement of a yaw degree value of -5 degrees will be considered as 'left' while a measurement of -20 degrees will be considered as 'extreme left'. A triangular member function is used to map the crisp in-

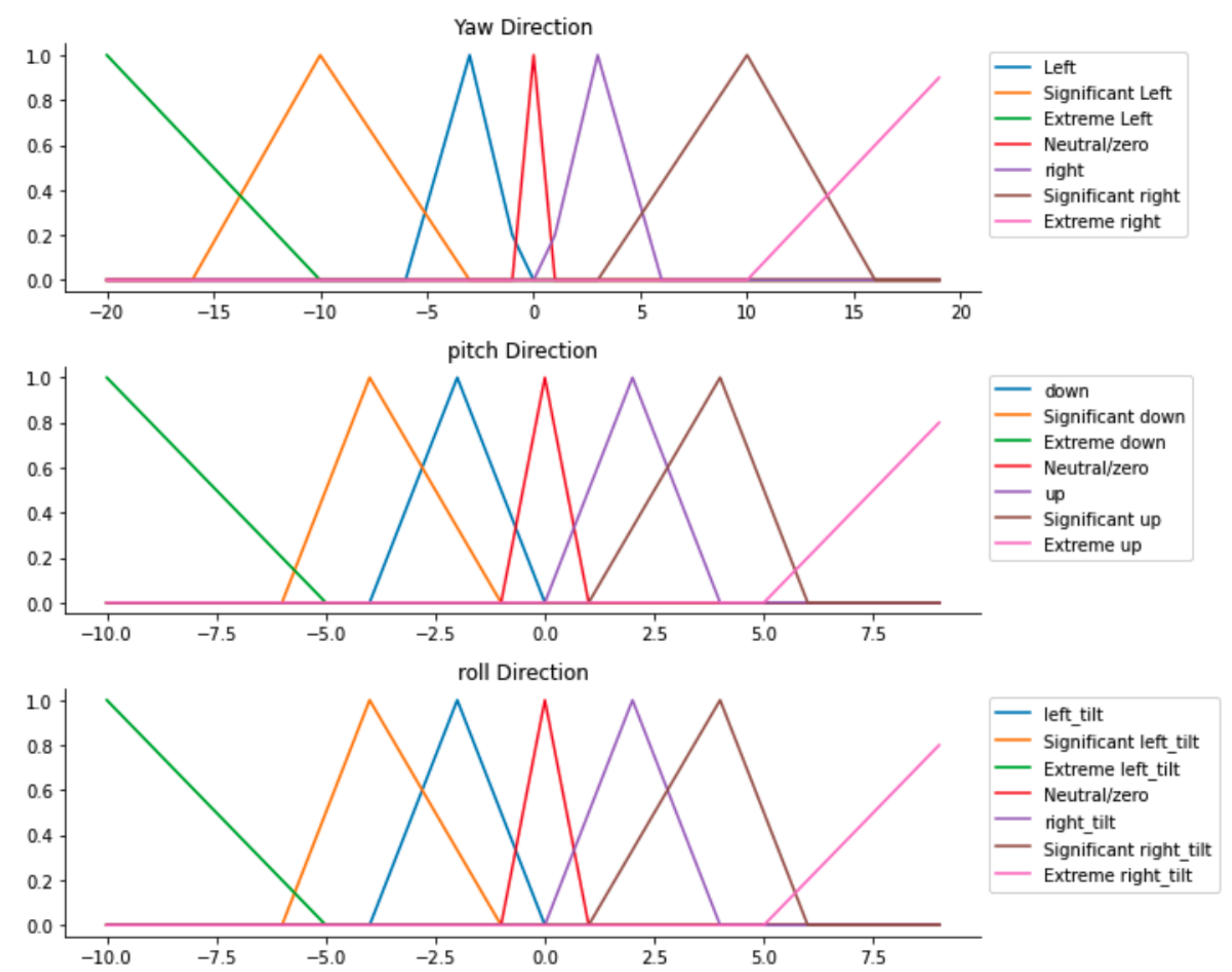

Figure 4.17: Fuzzy membership function used in fuzzification of Euler angles

put of Euler angles to a fuzzy set. This membership function is defined using three parameters base1,base2, and height. These parameters are represented on the x-axis and the corresponding fuzzy value is represented on the y-axis. According to (68), In general, a triangular membership function for a fuzzy set of length 1 is defined as follows:

$$\mu(x) = \begin{cases} 1, & \text{if } x = b \\ 0, & \text{if } x < a \text{ or } x > c \\ \text{(x-a)/(b-a)}, & \text{if } a \leq x \leq b \end{cases}$$

where x is the given input, a and c are the base of the triangular function, and b is the height of the triangle. This idea is adapted to define the triangular membership function for the given system of Euler angles and the membership graph is visualized in the figure 4.17.

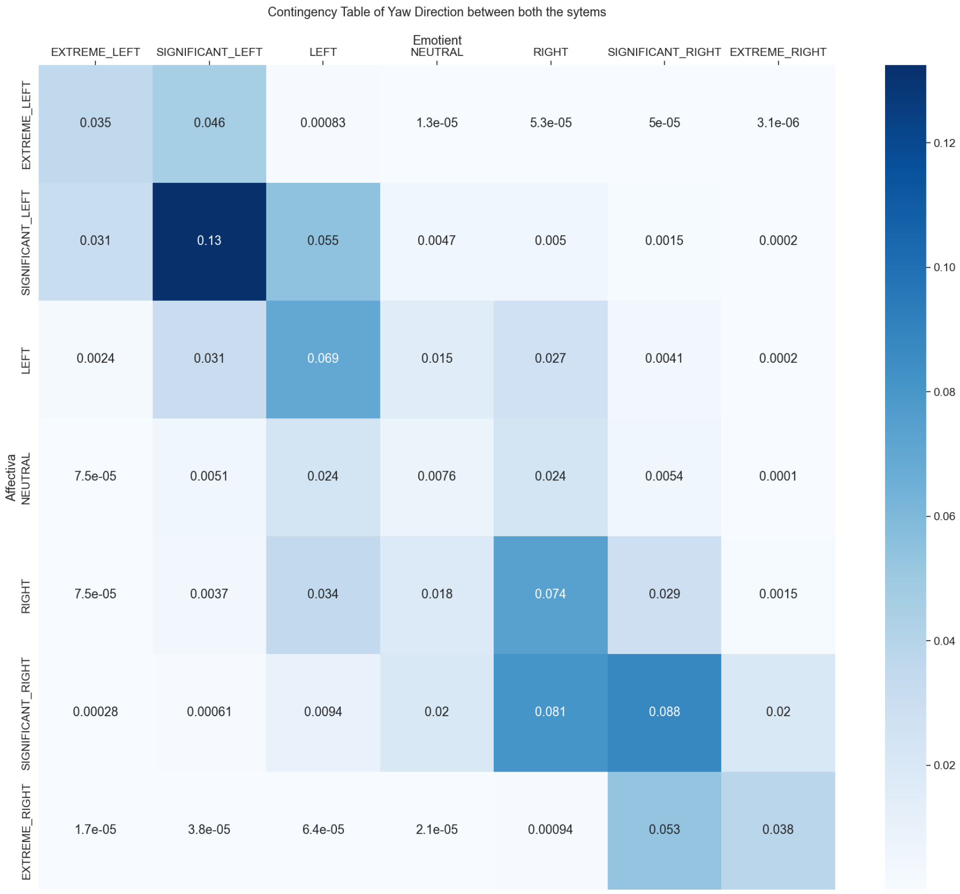

Figure 4.18: Agreement of head direction in the horizontal axis between Emotient and Affectiva

The yaw degrees in the range of -20 to 20 are mapped to seven discrete head directions. The negative yaw degree which indicates the head is to the left is further divided into significant left and extreme left. If the yaw degree is in a small range,i.e, between -5 and 5, it is considered neutral. Similarly, the head direction in the vertical axis is divided into 7 directions and is determined using pitch degrees. This triangular membership function ensures mapping of real-valued Euler angle to a discrete head direction eliminating crisp cut-off.

The estimated Euler angles from both the systems are fuzzified and the agreement between the systems is evaluated. A contingency table is created to evaluate the agreement of the systems in each of the 7 directions in the axis of interest. The main diagonal of the contingency table provides information on the percentage of frames where the systems are in complete agreement and the distribution of disagreement can be inferred from the other cells. The contingency table of head direction in the horizontal axis between Emotient and Affectiva is presented in the figure 4.18. Both systems seem to have a strong agreement on the head direction and the high values are concentrated on the main diagonal. Each value indicates the percentage of frames where the systems agree. For example, there are 3.5% of the entire frames where both the systems estimated the head direction is extreme left and 13% of the frames where both the systems estimated the head direction to be significant left. We can observe that there is a high concentration of frames on the same side of the head direction and the percentage of frames with contradicting head directions are negligibly small. For instance, In only 0.0002% of the frames, the head direction was estimated as significant left in Affectiva and extreme right in Emotient. To evaluate the statistical significance of the agreement, the

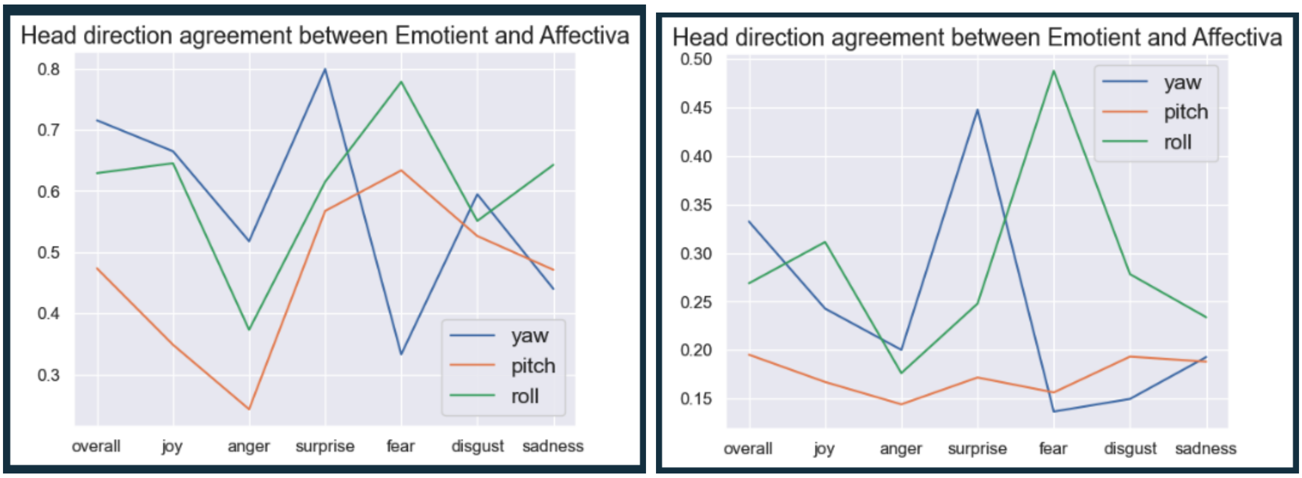

Figure 4.19: Comparison of variation of Cohen-kappa agreement before and after fuzzification

cohen kappa score is validated for the overall estimations of head direction from Emotient and Affectiva. The variation of the cohen-kappa score is visualized (figure 4.19)and compared with the previous approach.

Further, a similar bubble chart analysis for exploring head orientation and emotion is carried out with the fuzzified Euler angles.

	yaw_direction	pitch_direction	#frames	anger	joy	fear	disgust	sadness	surprise
0	SIGNIFICANT_RIGHT	DOWN	32970	1435 (4.4%)	805 (2.4%)	16 (0.0%)	359 (1.1%)	221 (0.7%)	977 (3.0%)
1	SIGNIFICANT_RIGHT	NEUTRAL	23147	1361 (5.9%)	1143 (4.9%)	3 (0.0%)	273 (1.2%)	91 (0.4%)	554 (2.4%)
2	SIGNIFICANT_RIGHT	UP	29787	1885 (6.3%)	884 (3.0%)	1 (0.0%)	289 (1.0%)	89 (0.3%)	609 (2.0%)
3	SIGNIFICANT_RIGHT	SIGNIFICANT_DOWN	36621	587 (1.6%)	358 (1.0%)	22 (0.1%)	429 (1.2%)	229 (0.6%)	926 (2.5%)
4	SIGNIFICANT_RIGHT	EXTREME_DOWN	24606	59 (0.2%)	223 (0.9%)	8 (0.0%)	442 (1.8%)	189 (0.8%)	717 (2.9%)
5	SIGNIFICANT_RIGHT	SIGNIFICANT_UP	40944	1153 (2.8%)	704 (1.7%)	2 (0.0%)	265 (0.6%)	60 (0.1%)	489 (1.2%)
6	SIGNIFICANT_RIGHT	EXTREME_UP	21836	368 (1.7%)	497 (2.3%)	1 (0.0%)	255 (1.2%)	59 (0.3%)	163 (0.7%)

	yaw_direction	pitch_direction	#frames	anger	joy	fear	disgust	sadness	surprise
0	SIGNIFICANT_RIGHT	DOWN	9313	168 (1.8%)	214 (2.3%)	639 (6.9%)	2 (0.0%)	311 (3.3%)	64 (0.7%)
1	SIGNIFICANT_RIGHT	NEUTRAL	8286	158 (1.9%)	276 (3.3%)	422 (5.1%)	10 (0.1%)	196 (2.4%)	55 (0.7%)
2	SIGNIFICANT_RIGHT	UP	17963	395 (2.2%)	688 (3.8%)	752 (4.2%)	21 (0.1%)	365 (2.0%)	109 (0.6%)
3	SIGNIFICANT_RIGHT	SIGNIFICANT_DOWN	6639	72 (1.1%)	108 (1.6%)	529 (8.0%)	1 (0.0%)	242 (3.6%)	46 (0.7%)
4	SIGNIFICANT_RIGHT	EXTREME_DOWN	4114	55 (1.3%)	15 (0.4%)	284 (6.9%)	0 (0.0%)	125 (3.0%)	34 (0.8%)
5	SIGNIFICANT_RIGHT	SIGNIFICANT_UP	32111	650 (2.0%)	867 (2.7%)	773 (2.4%)	0 (0.0%)	483 (1.5%)	175 (0.5%)
6	SIGNIFICANT_RIGHT	EXTREME_UP	95459	2559 (2.7%)	1639 (1.7%)	702 (0.7%)	11 (0.0%)	917 (1.0%)	472 (0.5%)

Figure 4.20: Comparison of relative frequency after fuzzification of Euler angles

4.20 presents the relative frequency of frames in different head orientations with respect to different emotions. It is to be noted that only frames where the yaw direction is 'significant right' is represented in the table. To represent the variation of head orientation with respect to different emotions, the bubble chart analysis is used and presented in figure 4.21.

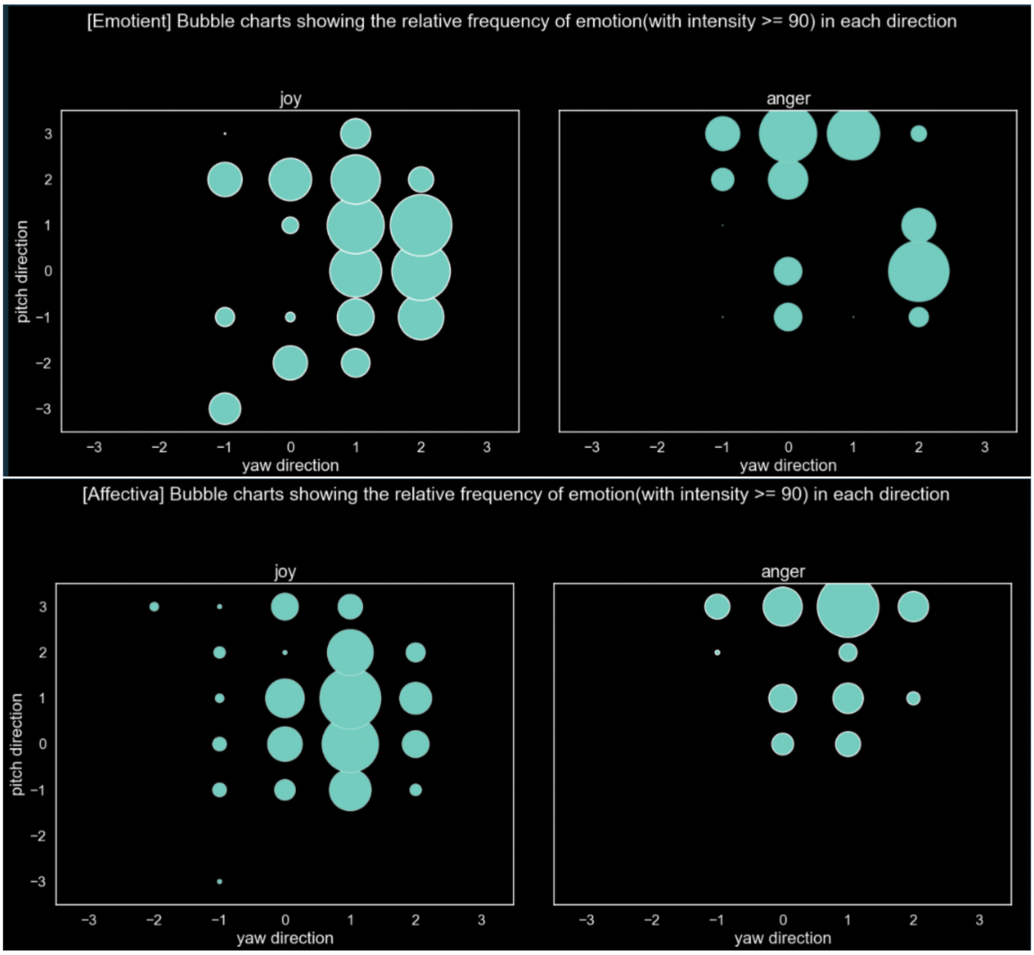

Figure 4.21: Comparison of the distribution of head direction after fuzzification of Euler angles

We can observe that the results obtained after fuzzification of Euler angles are in line with the inferences made from the previous analysis. However, the agreement between the systems evaluated using the cohen-kappa score has reduced significantly. Since, both the systems are not fully understood, this result provides a foundation to explore further in this direction.

4.3 Summary

This chapter presented various experiments and hypotheses developed to investigate the research questions. A systematic and statistical difference was identified between the systems for data distribution across professions. Further, the relationship between head orientations and emotions was established and the agreement between the systems was evaluated. The next chapter summarises the results and inferences derived from the analysis.

Chapter 5

Conclusion, Challenges and Future Work

5.1 Conclusion

The presented work added the profession of individuals as a new dimension to the precursor study (47) and investigated the systematic and statistical difference between two automatic emotion recognition systems. The results from this analysis indicate that there is a systematic difference between the systems(inter-system). Interestingly, the evidence for identified differences was evident within the systems as well(intra-system).

The major objective of the presented work was to explore a new modality and its influence on the perception of emotions. The relationship between head orientation and emotions was explored. The discussions and arguments presented in Chapter 4 explore whether the estimation of the head orientation is independent of the automatic emotion recognition systems. The mean of the estimated Euler angles are compared using a t-test and their distribution is compared in detail. Further, the variation of estimates with respect to emotions is explored and the results indicate that there is a significant difference in the estimated head orientation for the same set of videos by Emotient and Affectiva. Thus, one can infer that the computation of head orientation is not independent of automatic emotion recognition systems.

Head direction is the smallest correlate of head orientation and the relationship between head direction and emotions is analyzed. The agreement of the systems in estimating the head direction is estimated and evaluated using the cohen-kappa score. The results indicate that there is a strong agreement between the systems in estimating the head direction in all three dimensions(yaw, pitch, roll) and the variation of this agreement with respect to different emotions is presented. The agreement is highest for surprise and lowest for sadness. Head direction abstracts a lot of information about head orientation in the three-dimensional space. So the analysis is further taken to a granular level where, the interaction of head position with respect to different ranges of Euler angles is compared and presented as a bubble chart. The rationale behind this analysis is to understand if there is a specific head orientation associated with different emotions and to check the agreement of the systems. The result indicates that there is a clear association of head pointed up(positive pitch angle) with joy. However, comparing the different ranges of yaw and pitch angles brings in a crisp cutoff. The presented work attempts to tackle this issue using fuzzy logic. Seven discrete head directions are derived from the fuzzified Euler angles and the relationship of head direction and head orientation with emotions is evaluated. If the automatic emotion recognition systems are capable of understanding the relationship between head position and emotions, the application scope would be limitless. The presented work is an initial attempt to indicate that there exists a relationship, however, the relationship does not seem to be universal and is system-dependent.

5.2 Challenges

The major challenges faced during the research period are summarised below. The major challenge was the maintenance of data. A lot of effort was put into creating a seamless data collaboration environment. The major technical challenge was the inadequate availability of databases that include the ground truth for head orientation. For other modalities like facial expression, speech, and transcription, databases with the clear ground were widely available. Multiple efforts were taken to gather available datasets from researchers(email communications attached after appendix). However, the results were not fruitful. This aspect can be incorporated into future research and the inferences made from this study can be compared with the inferences obtained from analyzing human encoded head orientation database. Another challenge was the inconsistencies in the iMotion platform which did not allow us to create a standard preprocessing pipeline. For instance, certain videos have to be zoomed further to get a proper output from Emotient and Affectiva. Thus, many variations of preprocessing pipeline were experimented in a trial and error manner to ensure that the selected videos got processed. As a note to a future scholar who is interested in the extension of the presented work, these mentioned challenges can be tackled by optimizing preprocessing pipeline and obtaining a ground truth database for a comparative study.

5.3 Future Work

A careful examination of research methods was evaluated at the start of the research project to ensure that the presented work laid the necessary foundation for future exploration. The analysis of head orientation using fuzzy logic can further be explored to unveil the system behaviour which would facilitate a comparative study. The angular velocity of the movement of the head can be computed from the Euler angles and correlated with emotions. The possibility of representing the head orientation as a single measure instead of three angles can be explored. The representation of Euler angles as quaternions has been explored in solving the problem of head pose estimation. Similarly, pose angle sum (PAS) can be computed which is the sum of absolute values of Euler angles. These measures would act as a proxy measure of head orientation which can then be correlated to different emotions.

Bibliography

[1] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

[2] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological science in the public interest*, vol. 20, no. 1, pp. 1–68, 2019.

[3] H. G. Wallbott, "Bodily expression of emotion," *European journal of social psychology*, vol. 28, no. 6, pp. 879–896, 1998.

[4] A. Fernández, R. Usamentiaga, J. L. Carús, and R. Casado, "Driver distraction using visual-based sensors and algorithms," *Sensors*, vol. 16, no. 11, p. 1805, 2016.

[5] M. S. Bartlett, G. Littlewort-Ford, J. Movellan, I. Fasel, and M. Frank, "Automated facial action coding system," Aug. 5 2014. US Patent 8,798,374.

[6] J. Whitehill and J. R. Movellan, "A discriminative approach to frame-by-frame head pose tracking," in *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1–7, IEEE, 2008.

[7] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (cert)," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 298–305, IEEE, 2011.

[8] M. Bishay, K. Preston, M. Strafuss, G. Page, J. Turcot, and M. Mavadati, "Affdex 2.0: A real-time facial expression analysis toolkit," *arXiv preprint arXiv:2202.12059*, 2022.

[9] K. M. Yatheendra Pravan, C. Vogel, and K. Ahmad, "Head orientation of public speakers: Variation with emotion, profession and age (submitted, unpublished)," in *Future of Information and Communication Conference*, 2022.

[10] D. Phutela, "The importance of non-verbal communication," *IUP Journal of Soft Skills*, vol. 9, no. 4, p. 43, 2015.

[11] G. Giannakakis, D. Manousos, V. Chaniotakis, and M. Tsiknakis, "Evaluation of head pose features for stress detection and classification," in *2018 IEEE EMBS international conference on biomedical & health informatics (BHI)*, pp. 406–409, IEEE, 2018.

[12] S. Alghowinem, R. Goecke, M. Wagner, G. Parkerx, and M. Breakspear, "Head pose and movement analysis as an indicator of depression," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 283–288, IEEE, 2013.

[13] C. Sobin and M. Alpert, "Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy," *Journal of psycholinguistic research*, vol. 28, no. 4, pp. 347–365, 1999.

[14] N. S. Fox, "Clapping hands as a gesture of anguish and anger in mesopotamia and in israel," *Journal of the Ancient Near Eastern Society*, vol. 23, no. 1, p. 2401, 1995.

[15] C. E. Izard, "Facial expressions and the regulation of emotions.," *Journal of personality and social psychology*, vol. 58, no. 3, p. 487, 1990.

[16] P. Ekman, "Darwin's contributions to our understanding of emotional expressions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3449–3451, 2009.

[17] V. Sethu, J. Epps, and E. Ambikairajah, "Speech based emotion recognition," in *Speech and Audio Processing for Coding, Enhancement and Recognition*, pp. 197–228, Springer, 2015.

[18] A. Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon, "Emotion recognition in the wild challenge 2013," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, pp. 509–516, 2013.

[19] C. Trad, H. Hajj, W. El-Hajj, and F. Al-Jamil, "Facial action unit and emotion recognition with head pose variations," in *International Conference on Advanced Data Mining and Applications*, pp. 383–394, Springer, 2012.

[20] L. Zhang, L. Sun, L. Yu, X. Dong, J. Chen, and W. Cai, "Ieee transactions on biometrics, behavior, and identity science," 2021.

[21] J. N. Bassili, "Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face.," *Journal of personality and social psychology*, vol. 37, no. 11, p. 2049, 1979.

[22] P. Ekman, *Darwin and facial expression: A century of research in review.* Ishk, 2006.

[23] C.-H. Hjortsjö, *Man's face and mimic language.* Studentlitteratur, 1969.

[24] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pp. 3723–3726, 2016.

[25] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," *science*, vol. 221, no. 4616, pp. 1208–1210, 1983.

[26] R. E. Jack, W. Sun, I. Delis, O. G. Garrod, and P. G. Schyns, "Four not six: Revealing culturally common facial expressions of emotion.," *Journal of Experimental Psychology: General*, vol. 145, no. 6, p. 708, 2016.

[27] S. Gu, F. Wang, T. Yuan, B. Guo, and J. H. Huang, "Differentiation of primary emotions through neuromodulators: review of literature," *International Journal of Neurology Research*, vol. 1, no. 2, pp. 43–50, 2015.

[28] Z. Zheng, S. Gu, Y. Lei, S. Lu, W. Wang, Y. Li, and F. Wang, "Safety needs mediate stressful events induced mental disorders," *Neural plasticity*, vol. 2016, 2016.

[29] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *Proceedings of the national academy of sciences*, vol. 111, no. 15, pp. E1454–E1462, 2014.

[30] N. Alvarado and K. A. Jameson, "Varieties of anger: The relation between emotion terms and components of anger expressions," *Motivation and emotion*, vol. 26, no. 2, pp. 153–182, 2002.

[31] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Toward practical smile detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 11, pp. 2106–2111, 2009.

[32] D. Matsumoto, D. Keltner, M. N. Shiota, M. O'Sullivan, and M. Frank, "Facial expressions of emotion.," 2008.

[33] D. T. Cordaro, R. Sun, D. Keltner, S. Kamble, N. Huddar, and G. McNeil, "Universals and cultural variations in 22 emotional expressions across five cultures.," *Emotion*, vol. 18, no. 1, p. 75, 2018.

[34] D. Keltner, D. Sauter, J. Tracy, and A. Cowen, "Emotional expression: Advances in basic emotion theory," *Journal of nonverbal behavior*, vol. 43, no. 2, pp. 133–160, 2019.

[35] S. Namba, S. Makihara, R. S. Kabir, M. Miyatani, and T. Nakao, "Spontaneous facial expressions are different from posed facial expressions: Morphological properties and dynamic sequences," *Current Psychology*, vol. 36, no. 3, pp. 593–605, 2017.

[36] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 868–875, IEEE, 2011.

[37] K. D. Craig, S. A. Hyde, and C. J. Patrick, "Genuine, suppressed and faked facial behavior during exacerbation of chronic low back pain," *Pain*, vol. 46, no. 2, pp. 161–171, 1991.

[38] M. Heller and V. Haynal, "The faces of suicidal depression (translation). les visages de la depression de suicide," *Kahiers Psychiatriques Genevois (Medecine et Hygiene Editors)*, vol. 16, pp. 107–117, 1994.

[39] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, 1999.

[40] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 ieee computer society conference on computer vision and pattern recognition-workshops*, pp. 94–101, IEEE, 2010.

[41] M. Frank, M. Bartlett, and J. Movellan, "The m3 database of spontaneous emotion expression (university of buffalo)," *In pres*, 2010.

[42] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard, "Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database," *Image and Vision Computing*, vol. 32, no. 10, pp. 692–706, 2014.

[43] A. Tcherkassof and D. Dupré, "The emotion–facial expression link: evidence from human and automatic expression recognition," *Psychological Research*, vol. 85, no. 8, pp. 2954–2969, 2021.

[44] D. Dupré, E. G. Krumhuber, D. Küster, and G. J. McKeown, "A performance comparison of eight commercially available automatic classifiers for facial affect recognition," *PloS one*, vol. 15, no. 4, p. e0231968, 2020.

[45] S. Stöckli, M. Schulte-Mecklenbeck, S. Borer, and A. C. Samson, "Facial expression analysis with affdex and facet: A validation study," *Behavior research methods*, vol. 50, no. 4, pp. 1446–1460, 2018.

[46] E. G. Krumhuber, D. Küster, S. Namba, D. Shah, and M. G. Calvo, "Emotion recognition from posed and spontaneous dynamic expressions: Human observers versus machine analysis.," *Emotion*, vol. 21, no. 2, p. 447, 2021.

[47] K. Ahmad, S. Wang, C. Vogel, P. Jain, O. O'Neill, and B. H. Sufi, "Comparing the performance of facial emotion recognition systems on real-life videos: Gender, ethnicity and age," in *Proceedings of the Future Technologies Conference*, pp. 193–210, Springer, 2021.

[48] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 4, pp. 607–626, 2008.

[49] I. Chamveha, Y. Sugano, D. Sugimura, T. Siriteerakul, T. Okabe, Y. Sato, and A. Sugimoto, "Head direction estimation from low resolution images with scene adaptation," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1502–1511, 2013.

[50] F. Shojaiee and Y. Baleghi, "Pedestrian head direction estimation using weight generation function for fusion of visible and thermal feature vectors," *Optik*, vol. 254, p. 168688, 2022.

[51] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," in *CVPR 2011*, pp. 617–624, IEEE, 2011.

[52] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi, "Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation," in *2007 IEEE intelligent transportation systems conference*, pp. 709–714, IEEE, 2007.

[53] J. L. Tracy and R. W. Robins, "Show your pride: Evidence for a discrete emotion expression," *Psychological science*, vol. 15, no. 3, pp. 194–197, 2004.

[54] J. L. Tracy and R. W. Robins, "The prototypical pride expression: development of a nonverbal behavior coding system.," *Emotion*, vol. 7, no. 4, p. 789, 2007.

[55] B. App, D. N. McIntosh, C. L. Reed, and M. J. Hertenstein, "Nonverbal channel use in communication of emotion: how may depend on why.," *Emotion*, vol. 11, no. 3, p. 603, 2011.

[56] H. Toscano, T. W. Schubert, and S. R. Giessner, "Eye gaze and head posture jointly influence judgments of dominance, physical strength, and anger," *Journal of Nonverbal Behavior*, vol. 42, no. 3, pp. 285–309, 2018.

[57] M. Koppensteiner and K. Grammer, "Motion patterns in political speech and their influence on personality ratings," *Journal of Research in Personality*, vol. 44, no. 3, pp. 374–379, 2010.

[58] P. Werner, A. Al-Hamadi, K. Limbrecht-Ecklundt, S. Walter, and H. C. Traue, "Head movements and postures as pain behavior," *PloS one*, vol. 13, no. 2, p. e0192767, 2018.

[59] D. O. Johnson and R. H. Cuijpers, "Investigating the effect of a humanoid robot's head position on imitating human emotions," *International Journal of Social Robotics*, vol. 11, no. 1, pp. 65–74, 2019.

[60] S. Wu, Z. Du, W. Li, D. Huang, and Y. Wang, "Continuous emotion recognition in videos by fusing facial expression, head pose and eye gaze," in *2019 International Conference on Multimodal Interaction*, pp. 40–48, 2019.

[61] N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture.," *Emotion*, vol. 12, no. 5, p. 1085, 2012.

[62] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf, and L. M. Williams, "A public database of immersive vr videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures," *Frontiers in psychology*, vol. 8, p. 2116, 2017.

[63] M. Beringer, F. Spohn, A. Hildebrandt, J. Wacker, and G. Recio, "Reliability and validity of machine vision for the assessment of facial expressions," *Cognitive Systems Research*, vol. 56, pp. 119–132, 2019.

[64] P. Royston, "Approximating the shapiro-wilk w-test for non-normality," *Statistics and computing*, vol. 2, no. 3, pp. 117–119, 1992.

[65] N. M. Razali, Y. B. Wah, *et al.*, "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests," *Journal of statistical modeling and analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[66] S. Kokoska and D. Zwillinger, *CRC standard probability and statistics tables and formulae.* Crc Press, 2000.

[67] Z. Witkower, *The action unit imposter: head position influences social perceptions by changing the appearance of the face.* PhD thesis, University of British Columbia, 2021.

[68] W. Pedrycz, "Why triangular membership functions?," *Fuzzy sets and Systems*, vol. 64, no. 1, pp. 21–30, 1994.

Appendix

The following section is a supplement to analysis in the section 4 and continuation of table 4.5

Table 1: Results of T-test between different data gender populations

Gender	Euler Angle	T-statistic	pValue
MALE vs FEMALE (within emotient)	yaw	184.277714	0
	pitch	90.5161392	0
	roll	32.3769015	7.45E-230
MALE vs FEMALE (within affectiva)	yaw	82.1234878	0
	pitch	170.29196	0
	roll	9.75300946	1.79E-22
MALE	yaw	74.5099045	0
	pitch	-432.18938	0
	roll	72.2279069	0
FEMALE	yaw	13.2763817	3.21E-40
	pitch	-432.18938	0
	roll	72.2279069	0

The following section is a supplement to the regression analysis discussed in chapter 4, section 4.2.4.

```
                          OLS Regression Results
=======================================================================================
Dep. Variable:                    yaw   R-squared (uncentered):                   0.878
Model:                            OLS   Adj. R-squared (uncentered):              0.878
Method:                 Least Squares   F-statistic:                          2.048e+05
Date:                Fri, 19 Aug 2022   Prob (F-statistic):                        0.00
Time:                        00:36:33   Log-Likelihood:                          7628.1
No. Observations:               28450   AIC:                                 -1.525e+04
Df Residuals:                   28449   BIC:                                 -1.525e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
anger          0.0051   1.14e-05    452.523      0.000       0.005       0.005
==============================================================================
Omnibus:                      346.076   Durbin-Watson:                   0.071
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              358.673
Skew:                          -0.275   Prob(JB):                     1.30e-78
Kurtosis:                       3.006   Cond. No.                         1.00
==============================================================================
```

```
                          OLS Regression Results
=======================================================================================
Dep. Variable:                    yaw   R-squared (uncentered):                   0.890
Model:                            OLS   Adj. R-squared (uncentered):              0.890
Method:                 Least Squares   F-statistic:                          1.615e+05
Date:                Fri, 19 Aug 2022   Prob (F-statistic):                        0.00
Time:                        00:36:33   Log-Likelihood:                          5064.7
No. Observations:               20037   AIC:                                 -1.013e+04
Df Residuals:                   20036   BIC:                                 -1.012e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
joy            0.0054   1.35e-05    401.901      0.000       0.005       0.005
==============================================================================
Omnibus:                      154.829   Durbin-Watson:                   0.089
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              159.819
Skew:                          -0.206   Prob(JB):                     1.98e-35
Kurtosis:                       3.148   Cond. No.                         1.00
==============================================================================
```

Figure 1: OLS Regression Results

Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie>

Regarding Access to Warsaw Set of Emotional Facial Expression Pictures

3 messages

Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie> 30 June 2022 at 13:45
To: contact@emotional-face.org

Good Day,

I am Yatheendra Pravan. I am a researcher at Trinity College Dublin, Ireland.

My research is exploring the impact of head poses on the perception of emotions across cultures.

I am writing this email to request access to the Warsaw Set of Emotional Facial Expression Pictures. I tried filling out the request form, but I am unable to submit it.

This dataset would be of great help. Kindly consider my request.

Looking forward to hearing from you,

Regards,
Yatheendra Pravan K M

Michał Olszanowski <molszanowski@swps.edu.pl> 30 Ju
To: Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie>

Michał Olszanowski <molszanowski@swps.edu.pl> pt., 27 maj, 12:40
do Iris

Dear Yatheendra,

Attached please find a link to download the database:

https://www.dropbox.com/sh/jgw2moxd2cnpejq/AACFYWmQ6XnQkSI4nBA3E5t2a?dl=0

Good luck with your research!

Kind regards,
Michal Olszanowski

[Quoted text hidden]

--
Michal Olszanowski, Ph.D
SWPS University of Social Sciences & Humanities
Center for Research on Biological Basis of Social Behavior
Laboratory of Peripheral Psychophysiology
Chodakowska Street 19/31, PL - 03815 Warsaw
www.swps.pl, www.emotional-face.org

Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie> 30 June 2022 at 21:24
To: Michał Olszanowski <molszanowski@swps.edu.pl>

Thank you for giving me access to the dataset.

Regards,
Yatheendra Pravan K M
[Quoted text hidden]

Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie>

Request for Access to AMFED

1 message

Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie> 7 July 2022 at 20:16
To: amfed@affectiva.com

Good Day,

My name is Yatheendra Pravan. I am a research scholar at Trinity College Dublin.

My thesis is on understanding the impact of head poses on the perception of emotions across cultures.

I have attached the EULA herewith.

Looking forward to hearing from you,

Regards,
Yatheendra Pravan K M

AMFED_EULA.pdf
70K

Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie>

Requesting access to Transcultural Image dataset

Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie> 19 February 2022 at 17:07
To: jtejada@academico.ufs.br

Good Day,

I am Yatheendra Pravan K M. I am a researcher from Trinity College Dublin, Ireland. My research area is emotional intelligence and we concentrate towards non-verbal forms of communication.

I recently read your paper on "Building and validation of a set of facial expression images to detect emotions: a transcultural study". I am really fascinated by the extensive work done towards developing this dataset and interested in exploring further in this research direction.

I am writing this email to request you to consider sharing the dataset which will help me to progress towards my research goals.

Kindly let me know if you need any further information.

Looking forward to hearing from you.

Regards,
Yatheendra Pravan K M

Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie>

BP4D-Spontaneous: Binghamton-Pittsburgh 3D Dynamic Spontaneous Facial Expression Database

3 messages

Khurshid Ahmad <Khurshid.Ahmad@tcd.ie> 13 June 2022 at 06:10
To: "lijun@cs.binghamton.edu" <lijun@cs.binghamton.edu>
Cc: Yatheendra Pravan Kidambi Murali <KIDAMBIY@tcd.ie>

Dear Dr Lijun Yin
I am interested in the BP4D-Spontaneous data to study how head pose differs according to emotional states. The data will be used purely for academic use by me and my students, specifically Yathendra P. K. Murali who is doing a Masters Course in Data Science. He is looking at emotional leakage in politicians' videos in facial expressions and speech.

I will be grateful to you for allowing me access to the BP4D Spontaneous data base.

Many thanks

Best wishes
Khurshid Ahmad
Professor of Computer Science (1973),
School of Computer Science and Statistics,
Trinity College Dublin,
The University of Dublin
Dublin, Ireland

Lijun Yin <lijun@cs.binghamton.edu> 13 June 2022 at 22:55
Reply-To: lijun@cs.binghamton.edu
To: Khurshid Ahmad <Khurshid.Ahmad@tcd.ie>
Cc: "lijun@cs.binghamton.edu" <lijun@cs.binghamton.edu>, Yatheendra Pravan Kidambi Murali <kidambiy@tcd.ie>

Dear Dr. Ahmad,

Thank you for your interest in our BP4D database.

Before I send you the agreement form for signature, I would like to make a
note that due to a large amount of requests of our database, our
university
licensing office has changed a new policy to charge a modest handling fee
($500). Please let me know if it is ok with you.

If there is any problem, please feel free to let me know, I will do my
best to help you out.

Looking forward to hearing back from you.

Best,

Lijun.
[Quoted text hidden]
=-=
Lijun Yin, Ph.D. Office: Q18, Thomas J. Watson School of
Professor Engineering and Applied Science
Department of Computer Science Email: lijun@cs.binghamton.edu

Director, Graphics and Image Computing Laboratory
Co-Director, Seymour Kunis Media Core

Binghamton University URL: http://www.cs.binghamton.edu/~lijun
State University of New York Tel: (607)-777-5484
Binghamton, NY 13902 Fax: (607)-777-4729
=-=

Khurshid Ahmad <Khurshid.Ahmad@tcd.ie> 14 June 2022 at 06:24
To: "lijun@cs.binghamton.edu" <lijun@cs.binghamton.edu>
Cc: Yatheendra Pravan Kidambi Murali <KIDAMBIY@tcd.ie>

Dear Dr Lijun
Many thanks for your very prompt reply. Alas, I do not have access to such funds at the moment. The data was to be used by an MSc student and the next useful date for potential use of such data will be next year.

I will like to have the data if you can arrange. But I will understand if you cannot.

Best wishes

Khurshid Ahmad
Professor of Computer Science (1973),
School of Computer Science and Statistics,
Trinity College Dublin,
The University of Dublin
Dublin, Ireland

From: Lijun Yin <lijun@cs.binghamton.edu>
Sent: Monday 13 June 2022 18:25
To: Khurshid Ahmad <Khurshid.Ahmad@tcd.ie>
Cc: lijun@cs.binghamton.edu <lijun@cs.binghamton.edu>; Yatheendra Pravan Kidambi Murali <KIDAMBIY@tcd.ie>
Subject: Re: BP4D-Spontaneous: Binghamton-Pittsburgh 3D Dynamic Spontaneous Facial Expression Database

[Quoted text hidden]