12-2016

The Uncanny Valley The Effect of Removing Blendshapes from Facial Animation

Ameya Lonkar
axl1611@rit.edu

The Uncanny Valley
The Effect of Removing Blendshapes from Facial Animation

by

Ameya Lonkar

A Thesis Submitted
in
Partial Fulfillment of the
Requirements for the Degree of
Master of Science
in
Computer Science

Supervised by

Dr. Joe Geigel

Department of Computer Science

B. Thomas Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, New York

December 2016

The thesis “The Uncanny Valley - The Effect of Removing Blendshapes from Facial Animation” by Ameya Lonkar has been examined and approved by the following Examination Committee:

Dr. Joe Geigel
Associate Professor
Thesis Committee Chair

Dr. Reynold Bailey
Associate Professor

Prof. Warren Carithers
Associate Professor

Acknowledgments

Firstly, I would like to express my sincere gratitude to my and advisor, Dr. Joe Geigel, for his guidance and help, for his patience motivation and immense knowledge. I could not have asked for a better guide and mentor.

Besides my advisor, I would like to thank the rest of my thesis committee, Dr. Reynold Bailey and Prof. Warren Carithers, for their help and encouragement.

My thanks go to my friends Chinar Patil and Vishwanath Raman for their help and hardwork in setting up my thesis experiment. I could not have done this without them.

I am grateful to all the participants who took part in my study, without which this thesis would not be possible.

Last but not the least, I would like to thank my family and friends for their continued support throughout writing this thesis and life in general.

Abstract

The Uncanny Valley
The Effect of Removing Blendshapes from Facial Animation

Ameya Lonkar

Supervising Professor: Dr. Joe Geigel

Ever since its introduction in 1970 by the roboticist Dr. Mori, the uncanny valley has become an integral part of facial animation. While most of the work on the uncanny valley is focused on finding a way to surpass the uncanny valley and how certain things could be designed to increase or decrease the uncanniness of an animation, this paper investigates how removing certain blendshapes from the facial animation changes our perception about the eeriness of that animation. The goal is to identify the "necessary" blendshapes which bring about a significant change the perceived uncanniness of the animation. For achieving this, participants undergo visual tests which entails observing short animated clips expressing different emotions and recording the human-likeness and uncanniness, as perceived by the observers. The results validate some of the existing theories about the uncanny valley and also identify the necessary blendshapes for every emotion. The results show that removing the blendshapes not only affects the eeriness but also the perceived human-likeness and may change our perception of the emotion expressed in the animation.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The problem of the uncanny valley is very relevant to the field of computer animation as it can directly affect a success of an animated feature film or a video game. The animated feature films like The Polar Express [2] and Beowulf, where facial motion capture was used, did not have the desired effect on the audience. The audience could not empathize with the characters in that movie because of the uncanny valley. It can also be used for our advantage where we might want something to look uncanny(eg: zombies should look uncanny). This is the reason why the uncanny valley is being taken seriously by animators. Many researchers are trying to find a way to surpass or avoid the uncanny valley. This paper addresses the concept of the uncanny valley in the facial animation driven by blendshape interpolation [3]. Blendshape facial animation is technique based on weighted interpolation of static facial poses which act as keyframes for the animation. The objective is to find out how different blendshapes that are used for facial animation, affect the perceived eeriness of the animation. This thesis delves deeper into the concept of the uncanny valley to find out how different blendshapes that are used for facial animation, affect the perceived eeriness of the animation. For testing this, different animations for a 3D face model were created. These animations were created using different set of blendshapes which were then showed to the subjects and their response on how human-like and familiar an animation looks, was recorded. The observers were also asked to guess the emotion conveyed through every animation. Then based on the empirical evidence, conclusions were drawn regarding effect of removing the blendshapes from the animation. The set of blendshapes that will be used to create these animations were decided based on which parts of the face have the

most significant effect of the uncanny valley. This paper extends the previous work on the uncanny valley by considering things that significantly affect the uncanny valley. This will help in quickly identifying the key blendshapes that affect the perceived eeriness of the animation, simultaneously taking into consideration the reasons why the uncanny valley occurs. This will answer if all the blendshpes are important while animating a face and also help to identify the features of the face that invoke uneasiness and the ones that do not.

The experiment, attempts to determine the blendshapes that can be termed as "necessary" so that the animation does not go into the uncanny valley. As a result of this, the blendshapes that do not fall in the necessary category can be skipped, making real-time facial animation much more efficient and less uncanny.

Chapter 2

Background and Related Work

A vast amount of research has been conducted to find out what causes the uncanny valley and how different aspects of the animation can be modified to make the animation less or more uncanny (eg. we want zombies to be more uncanny). This paper will analyze previous work to form a hypothesis about the “necessary” blendshapes that play an important role to cause the eerie sensation.

2.1 The Uncanny Valley

The concept of “The Uncanny Valley” was initially introduced in robotics by Dr. Masahiro Mori [13] in 1970. His work stated that android robots made too similar to humans can fall into ”The Uncanny Valley” where a very high degree of human likeliness evokes an unpleasant feeling in the viewer. According to his theory, the perceived familiarity of a robot is directly proportional to human-likeness of the robot. This continues up until a certain point where robot is perceived as more strange than familiar. This sudden revulsion or negative feeling is the uncanny valley (Figure 1).

In recent years, the uncanny valley has become an important phenomenon in computer animation and virtual reality. It has been found out that if an animated character falls into the uncanny valley, it will evoke a sense of eeriness or revulsion in the viewer which will lead to the viewer not feeling any empathy towards the character. This will cause the animated feature film or an animated short or a video game to not have the desired effect on the viewer. This has been observed in feature films like The Polar Express, Beowulf,

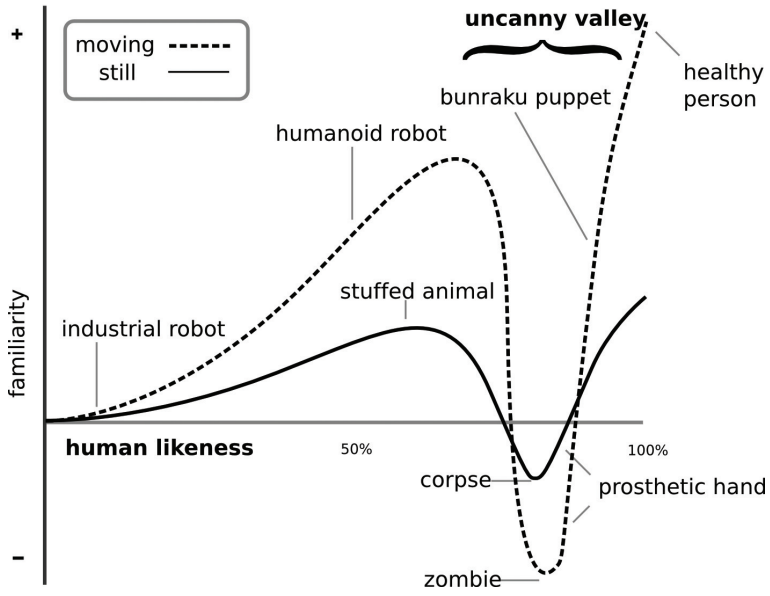

Figure 2.1: The Uncanny Valley as proposed by Mori and later translated by Macdorman [13]

etc. The most important reason these films did not have the desired effect on the viewers was the uncanny valley.

Extensive research in this area led to several theories regarding the uncanny valley which are based on what Dr. Mori said and vary slightly. Karl MacDorman, when he was revisiting the uncanny valley [9], where the observers rated still images of a robot morphing into a human on a 9-point scale, found out that the uncanny valley was around 35% to 40% of the graph. The uncanny valley was found to be much earlier than Mori's graph which has the valley around 80% to 85%. Also Tinwell and Grimshaw [20] when they conducted a similar test on a 9 point scale, they found the valley to be at around 50% to 55% (again much lower than Mori). However the major difference between these two experiments was that Tinwell and Grimshaw [20] included speech and movement as a part of the experiment, whereas Macdorman [9] only had a few animated shorts. The cause of this difference in empirical evidence as stated by Macdorman are the factors beyond appearance of the animated character (like movement, posture, speech, etc.) which can significantly affect the uncanny valley. Seyama [14] assesed the uncanny valley by recording the ratings for

static facial images, where human faces were created morphed from dolls, masks,etc. It was found that only abnormal features evoked the uncanny feeling.

2.2 Causes

The major causes for the animation to fall in the uncanny valley based on Macdorman and Ishiguro (2006) [11], Green and Macdorman (2008) [7], Donovann (2003) [5], Macdorman (2009) [10] and Dill et al. [4] are as follows:

2.2.1 Mortal Salience

Viewing an uncanny robot or an uncanny animated face evoke a subconscious feeling of mortality. Something that can be potentially harmful, that can replace us or annihilate us. This is very apparent in movies and games that involve zombies or undead creatures which scare us. This might one of the reason why the uncanny sensation is caused.

2.2.2 Pathogen Avoidance

The uncanny feeling may stem from a cognitive mechanism that evolved for pathogen avoidance from anything or anyone that looks unhealthy. Robots and animated faces might fall into this category because the more they resemble humans, the stronger the aversion we have to their defects.

2.2.3 Violation of Human Norms

If an entity look sufficiently non-human, the human-like features of that entity will be amplified. But on the other hand, when an entity resembles a human, the defects or the discrepancies will be more apparent. Even a slight deviation from human form will be easily noticeable. This might be the most significant reason for the uncanny valley.

There is empirical evidence that suggests many other causes for the uncanny. As seen, movement amplifies the graph of the uncanny valley. If the entity under consideration is

very human-like, but its movement does not match a human moves, then the object will appear stranger. Tinwell and Grimshaw [20] observed that sound plays a very important role in eeriness of the animation. The lack of lip synchronization enhances the uncanny sensation which is a combination of both sound and movement. They also found that there was a strong correlation between the appearance and the voice. That is, if the voice of a character actually sounded like that character and not out of place.

Sigmund Freud (1919) [6] states that the uncanny is the revelation of the repressed. Its not something that shouldn't be there but something that was there and is now visible. The eeriness occurred when the situation evoked a sinister revelation of what is normally concealed. Freud suggested that this revelation of repressed that should have remained hidden may not necessarily be an unfamiliar concept. It may be something that was once forgotten but we were suddenly alerted to it. Interestingly this hidden uncanniness may not only be in others but also what may be repressed and hidden in oneself. Freud summarized that the uncanny does invoke feelings of uncertainty, but primarily uncertainty in oneself.

Tinwell (2014) [17] further explains Freud's theory in her book. She suggests that the animated face with strange facial expressions may trigger doubt of who we are and how we are perceived by others. We as humans, tend to find similarity rather than differences in other humans as this helps better understanding and validation of oneself [17]. We try to find similarities between us and others to ensure smooth interaction with them and also a means of self-verification for the purpose of existential security and psychological coherence [15]. So, in mimicing others facial movements we not only get a better understanding of their emotional state but also establish a better understanding of our own identity [22]. Hence, we relate more to people who respond to us in a way that fits our self concept and feeds our ego. When we see an animated character with a highly human-like appearance that lacks the ability to mimic our facial expressions but also prevents us from understanding their emotional state, we might perceive that we are not being recognized as we do when we interact with other humans. This way, not only is self-verification not possible but this might also bruise our ego [17]. This might be one of the causes for the uncanny

valley.

Familiarity and effect of habituation to an animated character may change the perceived eeriness of the character. Minato [12] suggested that habituation may affect the interaction of human subject to an android. He also speculates if prolonged interaction with the android will cause the human subject to be more accepting of the android, although he cites that further tests may be required. On the other hand Tinwell and Grimshaw (2009) [18] also conducting tests on the same 9-point scale observed that familiarity and habituation do not play a significant role in reducing the uncanny valley. A character perceived as not human-like in the beginning will stay that way. Although, an uncanny face might become familiar over time, the eeriness will never go away. There were two types of test subjects involved in the experiment, ones who were accustomed to 3D modeling and animation and others who were not. It was observed that the average ratings from these observers were similar. This shows that familiarity with computer graphics or animation does not make a difference.

2.3 Uncanny Valley in Virtual Characters

Macdorman [10] and [9] show that there are several factors other than the face that influence the uncanny. More so, when considering the uncanny valley in virtual characters. Things beyond what Tinwell and Grimshaw state(lip synchronization, strong correlation betwwen the voice and the character, body movement, etc.) like the texture applied to the skin, the lighting of the scene, number of polygons used for rendering. According to [9] the increase in the number of polygons makes the face look less uncanny. The use of highly photorealistic texturing makes it look more human but, makes the uncanny all the more evident. Macdorman states that the uncanny valley is not just associated to a character being human-like. A non human-like character can look uncanny because of other factors like facial distortions and lack of movement in the facial expression.

Facial expression plays an important part in the perceived eeriness. Many of the test results [10], [19], [18], [8] and [16]indicate that lack of non verbal communication and facial expression, specifically in the upper facial region affect the perceived eeriness the

most. Hodgins et al. [8] tested this by creating three vignettes. These consisted of a two actors enacting an interaction between a couple in an apartment. In the first scene Milk, the woman is angry because there is no milk in her cereal. In the second scene Money, the man is angry because he discovers that the woman has bought a rather expensive piece of clothing. The third scene; Moving Out, shows that they have split up and the man arrives unexpectedly to take his belongings and the woman has a guest at home. Facial motion capture and eye tracking was used for simulating the expressions of the actors on the animated characters. To test how the lack of facial motion affects the uncanny valley, the different experimental conditions were the face missing some feature. They were Full (F): the unmodified animation, No Face (NF): the face and eyes were completely frozen, No Eyes (NE): the face was animated but the eyes did not move, Offset Eyes (OE): one eye was rotated by a constant amount, giving an effect similar to Amblyopia(or lazy eye), Low Face, Eyes (LF E): the motion of the jaw was restricted to one degree of freedom (open/close for speech synchronization), but the eyes were animated, Low Face, No Eyes (LF NE): face as with LF E, but the eyes were not animated, Half Face (HF): only half the face was animated(this condition was chosen to mimic Bells palsy), No Arm (NA): one arm was not animated and was held stiffly in front of the character as if in a sling, Noisy Arm (NyA): a tremor was added to the characters arm, No Pelvis (NP): the top half of the body only was animated by removing data from the root and legs, No Sound(NS): No sound was played. The test was split into two by recording short vignettes and long vignettes. The long ones just had combinations of face and sound NF, NS as the tests while the short vignettes had combinations of the other conditions. The participants were asked to rate anger and sympathy for the man and the woman. It was observed that the OE, HF and the NA conditions were least preferred by the observers. As these conditions exhibited signs of an unhealthy person. This further affirms Macdorman's theory of disease avoidance as one of the causes of the uncanny valley. These tests also indicated that facial anomalies were more significant and more disturbing than uncanny body movements. The NF - No Face condition was rated as most strange. Hodgins et al. state that this happens because no facial

expressions resemble a dead body. The frozen face was still rated as the angriest of all as it resembles a stony-face which is considered as the highest expression of anger. Absence of audio makes the woman look more angry in the Moving Out scene as it makes her body movements look more aggressive. Observations from [8] show us again that the uncanny valley is affected more by facial motion and expression. However, body movement and sound play an important role in the perceived uncanniness. In addition to this, Hodgins also considers attention as an important factor as attention guided towards the face or the body will make the anomalies in that feature more significant.
[8] also proposes that the uncanny valley is a multidimensional model as opposed to being a simple graph.

Tinwell et al. [19] also analyses facial expressions and their correlation with the uncanny valley. [19] tested this by making users rate facial expressions of 3 faces. the three faces are - an actual human, a fully animated face and a face which has upper facial expressions missing. The six basic facial expressions were used - anger, fear, disgust, happiness, sadness, surprise. The participants were all male and from a computer animation, 3d modeling, software background, so that they would be accustomed to the 3d virtual characters. They were asked to rate the level of human likeness and familiarity. An actor displaying the six emotions recorded the videos and for the animated character the actor recorded the six expressions using a facial motion capture device. Action units were used to control facial muscles. The virtual character with upper facial expressions missing, had the action units that controlled eye brows and eyelids were disabled. It was observed that the human as predicted, was the most familiar. This was followed by the fully animated face and the partially animated face without upper facial animation was the most strange. The facial expressions were most accurately predicted for the human, followed by the fully animated and then the partially animated face which had the lowest accuracy for recognizing facial expressions.
The facial expression also plays an important role in the perceived uncanniness. It was observed that fear and disgust are perceived to be the most uncanny. Tinwell et al. [19]

propose that fear and disgust makes are the emotions that are one of the reasons that cause the uncanny valley(mortal salience and disease avoidance). Surprise can sometimes be confused as fear, making ambiguity the reason why it was also perceived as uncanny. Happiness and anger are less of survival emotions as compared to fear, sadness and disgust. Also, the least facial movement is required to express anger and disgust making these emotions as most uncanny. Minimum movement of the face also resemble a dead person which could also be the reason of the perceived uncanniness. This paper proposes that the emotion exhibited by a virtual character is important because it helps predict their likely behavior. When the combination of voice and facial expressions of a person is used, the viewers expect some sort of congruence between the two. Even a minor incongruence causes unpredictability to factor in, which causes the strange feeling or eeriness. The recipient may perceive this unpredictable behavior as a potential threat. This makes the facial expression, especially the upper facial movement extremely crucial towards the perceived uncanniness. It is very important to control the upper facial movements when attempting to display emotions like sadness, disgust and fear.

Dill et al. [4] conducted several experiments to evaluate different aspects of the uncanny valley. The participants were showed 14 animated faces. Two separate tests were conducted on different set of participants, one was with stills of the 14 animated faces and the other included video clips of the 14 faces. A set of six questions were asked to every participant, they were - a) Was the character a real person or not? b) If CG generated how realistic did the person look? c) Do you know the character? d) Would you describe the character as sympathetic or not? e) Does the person appear strange? f) Which part made the person look strange?(eyes, mouth, nose, hair). For the first five questions the results of this experiment recreated a graph similar to Mori's uncanny valley. The final question was meant to investigate the part of the face that causes the uncanny valley. It was observed that for the still images eyes were recorded as the cause 35.25 % of the time and mouth was the cause 29.98% of the times. The major factors were eyes and mouth which contributed to 65.23 % of the total number of participants. For the videos, it was observed that eyes

were at 35.18% and mouth was at 37.85%. For the videos, eyes and mouth were found to be the most at 70 [4] shows that facial expressions play an important role in the perceived uncanniness. The results show that not only upper facial movements near the eyes, but also facial movement near the mouth is critical to the uncanny valley.

It is clear that uncanny valley is more complicated than a single graph with a lot of different factors affecting it. [10], [19], [18], [4], [8] and [16] suggest that the uncanny valley is a multidimensional model and it does not just depend on the animated character being human-like or not. Non-verbal communication, facial expressions comprising of various facial movements are detrimental to the perceived uncanniness.

Chapter 3

Design and Implementation

My solution consists of two steps. First is creating the stimuli for the five emotions and 9 sequences for every emotion. The stimuli were created using the facial motion capture technology of Faceshift [21] and the the preset photorealistic CG model 'Macaw' available with Faceshift was used for creating the animations. Faceshift requires the user to train its motion capture using 23 different expressions.

It can also be trained with fewer expressions, but this will decrease the tracking quality. A set of 51 blendshapes are used by faceshift for tracking. The value of every blendshape can be controlled and the effect causes on the animation can be adjusted. Some of these blendshapes will be disabled to create eight sequences for every emotion and there will be a "basic" sequence for every emotion which includes all the blendshapes. An actor recorded five sequences for the five emotions using faceshift and the primeSense sensor. The five emotions used are - fear, disgust, surprise, anger and happiness. The different blendshapes that were disabled to create different stimuli are - Mouth, Upper Lip, Lower Lip, Brows, Sneer, Both(Mouth and Brows), Puff and Cheek Squint. The reason for this is that fear and disgust are the emotions that are perceived as the most uncanny as these are survival based emotions. Surprise is easily confused with fear and therefore affects the perceived eeriness significantly. This will help evaluate the importance of each blendshape better. Happiness and anger are added to evaluate the amount of difference blendshapes can cause to the emotion happiness or anger which according to empirical evidence should be perceived to be the least uncanny. The second step is to show these animated sequences to the participants and ask them to answer a questionnaire, which will ask for ratings on a

Name	Description
Neutral	The actor's neutral expression with closed mouth. For best results the actor should not fully clench the teeth but leave some space between the teeth instead. Note though that the lips should be closed.
Open	The mouth should be opened widely without stretching the mouth corners.
Smile	A wide smile with closed lips.
BrowsDown	As for an angry expression, the brows should be put down as much as possible, while keeping the remainder of the face unchanged.
BrowsUp	As for a surprised expression, the brows should be pulled up, but the remainder of the face should stay as still as possible.
Sneer	Wrinkle the nose (as for a disgusted expression) and move the brows down, similar to the central part of the angry expression.
JawLeft	Move the jaw to the left, without opening the lips.
JawRight	Move the jaw to the right, without opening the lips.
JawFront	Move the jaw to the front, without opening the lips.
MouthLeft	Pull the mouth to the left.
MouthRight	Pull the mouth to the right.
Dimple	As in a smile, but without rasining the mouth corners.
ChinRaise	Move the mouth towards the nose.
Kiss	A kiss with the lips put together and pushed to the front.
Funnel	Similar to kiss, but with an open mouth.
Frown	As for an unhappy face with the mouth corners down.
M	Rolling the lips in, as it happens when pronouncing an exaggerated "m".
Puff	Blowing out the cheeks.
Chew	Open the jaw but keep the mouth closed.
MouthPress	Press your lips.
Stretch	Stretch your mouth corners as much as possible to the opposite sides. Different from dimple.
LipLowerDown	Lower your lower lip without opening the jaw or moving the mouth corners.
LipUpperUp	Raise your upper lip without opening the jaw or moving the mouth corners.

Figure 3.1: The different expressions for which the user should train faceshift

9-point scale for different questions. These sequences will be 3 to 4 seconds long and will be shown in a random order. To evaluate the data collected, the answers to the questions for all of the 8 stimulae with missing blendshapes was compared to the ”basic” stimuli for that emotion. These resuts were then aggregated to see which blendshapes affect the percevied eeriness the most and this will be done for every emotion.

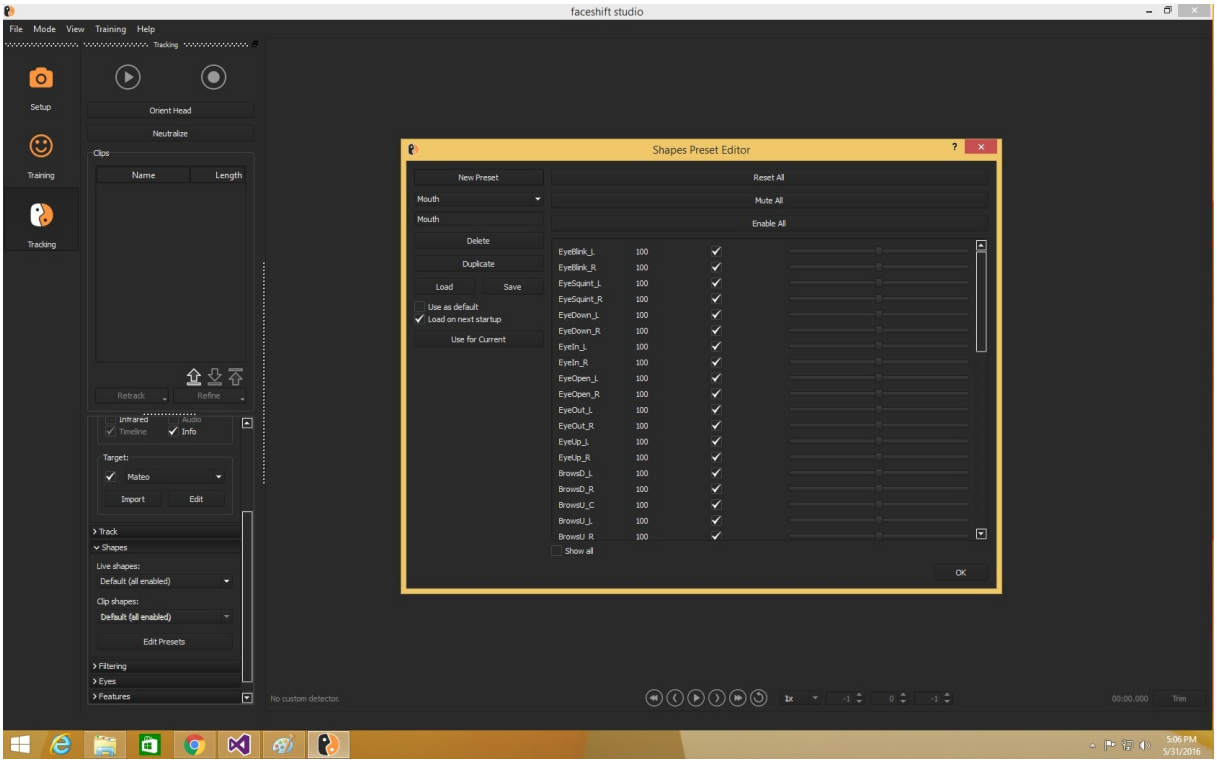

Figure 3.2: Some of blendshapes that can be disabled to create the 8 different variations of the video

3.1 The Experiment

The aim of this study is to test if adding or removing certain blendshapes while creating the animation causes a significant difference in the perceived uncanniness of the animation. This will be assessed by analysing the answers of the questionaire which will be answered by the participants after watching each video. All participants will be seeing the same set of videos in different random order.

There are 9 different video sequences created for each of the 5 emotions making the total video count 45. Each video is 3 to 4 seconds long. Along with human-likeness and perceived eeriness, the accuracy with which the participants can guess the emotion was also recorded. The reason that the accuracy of the emotion is so important is that, there might be a case where the removed blendshapes cause no significant difference in the perceived uncanniness but causes a different emotion to be potratyed. In this case, the blendshape

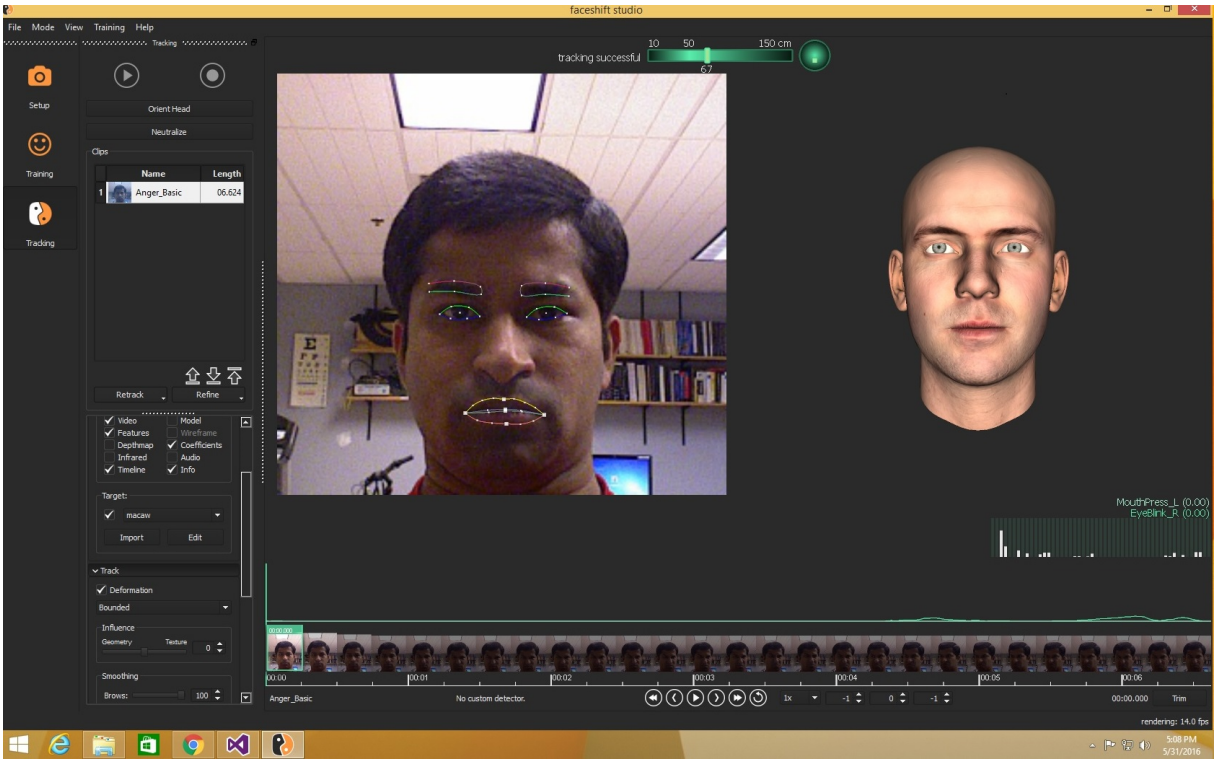

Figure 3.3: How a loaded video clip of the recorded facial expressions looks

that causes the wrong emotion to be conveyed are as bad as the blendshapes which cause the animation to look uncanny.

3.1.1 Participants

The participants for this study were volunteers from the Rochester Institute of Tecnonlogy gathered through vocal requests and through flyers posted in the Rochester Institute of Technology campus. Volunteers were informed of the experimental procedure and also given information of what exactly do the questions mean. They were also told that they may stop the experiment whenever thaey want to. Before the experiment they signed a consent form which describes the experimental procedure in detail and provides all information that meets the IRB standard. The IRB certification for this experiment was obtained before the experiment.

3.1.2 The Questions

The four questions which the participants were asked to answer for every video were -

Question 1 - ”How human-like do you find the character to be?” - This question refers to the character model which is the most human-like model that comes as a preset with faceshift called ’macaw’. This question covers the human-likeness of the animation. Participants are asked to rate the human-likeness of the character and not his expression or acting. However, the emotion and expression of the character could possibly affect how the participants answer this question. This question is answered on a 9-point scale.

Question 2 - ”How strange do you find the clip to be?” - This question refers to the entire animation and not just the character. This questions deals with the perceived eeriness of the animation. The rating for this question changes with the change in the missing blendshapes and also the emotion under consideration. This question is answered on a 9-point scale.

Question 3 - ”What do you think is the emotion in the clip?” - This question adresses the other important aspect of the experiment which is conveying the correct emotion, when

Table 3.1: The stimuli with the blendshapes removed and their descriptions

Basic/BL1	Contains all of the blendshapes
Mouth/BL2	The Smile movement was disabled
Brows/BL3	The Brow movement was disabled
Both/BL4	BL2 and BL3 were removed
Sneer/BL5	The sneer movement was disabled
UpperLip/BL6	The Upper Lip movement was diabled
LowerLip/BL7	The Lower Lip movement was disabled
Puff/BL8	The Puff movement was disabled
CheekSquint/BL9	The Cheek Squint movement was disabled

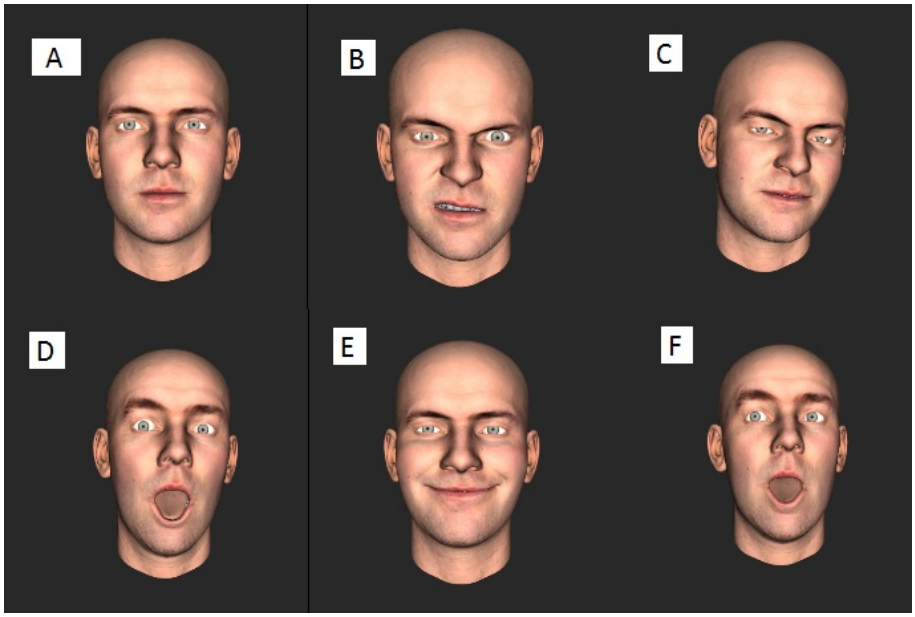

Figure 3.4: How the face looks when it is displaying differemt emotions where, A = Neutral, B = Anger, C = Disgust, D = Fear, E = Happiness and F = Surprise

the ”unecessary” blendshape is removed. The participants select one of the five options which are - fear, disgust, surprise, happiness and anger. This question is very important along with the human-likeness and eeriness as the possiblility of wrong perception of the emotion could be a case worse than the uncanny valley

Question 4 - ”How confident are you about you choice in question 3?” - When certain blendshapes are removed some of the videos will display a mix between 2 emotions and cause confusion. This question helps to evaluate the correctness of their choice. Also, it is easy to get confused between fear and surprise and in that case the removed blendshape may not have caused the confusion but the fact that fear and surprise are easy to be percived as eachother.

Chapter 4

Analysis

4.1 Hypothesis

The goal of this experiment is to determine the “necessary” blendshapes that minimize the uncanniness of the resultant animation and identify the blendshapes which heighten the eeriness of the animation. Based on the empirical evidence gathered from the related work, the hypothesis underlying this thesis was that the blendshapes that are essential for smooth interpolation of certain facial expressions like fear and disgust will be the ”necessary” blendshapes. Especially all the blendshapes that have change in the upper facial region(eyes, eye brows) and mouth. So blendshapes like mouth-open, eyebrows-up, eyebrows-down, sneer, upper lip and lower lip will be the “necessary” ones and blendshapes like puffy-face or smile will not be termed as necessary. The reason for this is that the expressions that we termed as necessary are the ones which will have the most effect on conveying the emotions of fear, disgust and surprise. The human-likeness ratings should not vary for the same user, as the 3D model used is the same.It might vary from user to user as different users can have a different perception of what they consider to be human-like.

4.2 Results and Discussion

There are three main aspects of this experiment, the perceived emotion, human-likeness and the eeriness. The results are discussed in reference to the Basic video, as it is the stimuli with all the blendshapes present. So it is expected to be the most human-like, least eerie and with the highest correct prediction percentage.

4.2.1 Emotion

Fig. 3 shows the correct prediction percentage for Anger Disgust and Fear and Fig. 4 shows data for Happiness and Surprise. The correct prediction percentage for Disgust and Fear for the Basic stimuli was lower than the other emotions. Anger had high prediction accuracy even when the blendshapes were removed. Only Sneer, UpperLip and LowerLip decreased the prediction accuracy when removed. The sneer and pressing the upper lip and lower lip together results in to the angry expression and removing these blendshapes might have made the animation to be perceived as some other emotion. Disgust had low prediction accuracy for the Basic video. The removal of Mouth and Brows caused the accuracy to increase by more than 10%. Removing Sneer, UpperLip and LowerLip decreased the accuracy drastically. The nose wrinkiling action and the curling of upper lip are used to express disgust. Removing these blendshapes changes the perceived emotion possibly to anger. However the removal of Mouth and Brows improved the prediction accuracy. This might have more to do with the limitations of the motion capture and animation software and the acting for the motion capture. The suppressed movement of the edges of the mouth and brows might have been perceived as more disgusted than the Basic. For Fear, removing every blendshape resulted into increased prediction accuracy. The most significant increase was when both Brows and Mouth(BL4) were removed, which resulted into a 20% increase. Removing UpperLip, LowerLip, Puff and CheekSquint resulted into the accuracy being increased by 10% to 15%. The removal of Mouth+Brows might have made the face to appear as rigid and without movement which would make it look more terrifying. For Happiness, removing the Mouth blendshape resulted into decreased prediction accuracy, when it dropped by 40%. The Mouth blendshape controls the edges of the mouth and by removing it the smile will look like a frown with the lips pressed together. Surprise had a high prediction rate. Even when the blendshapes were removed it still had 70% accuracy in the worst case.

The prediction confidence ratings were generally high and did not exhibit a trend with changing emotions and removing blendshapes.

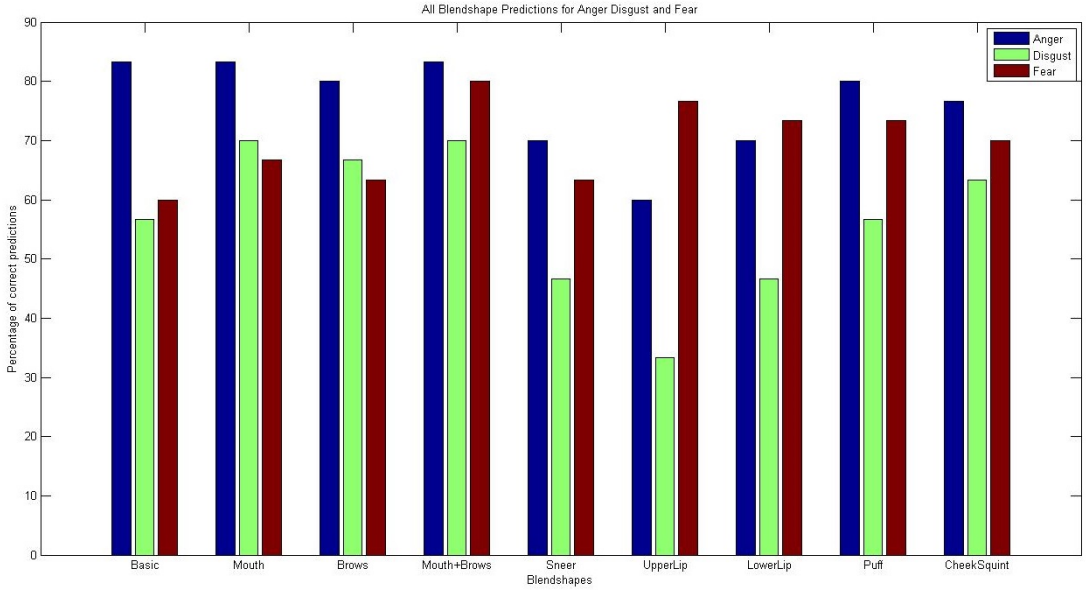

Figure 4.1: The Blendshape-wise Correct Prediction Percentages for Anger,Disgust and Fear

4.2.2 The Uncanny

Fig. 5 and Fig. 6 show the average eeriness ratings with error bars for all emotions. The average ratings for familiarity varied accross the different emotions. The average ratings for the Basic stimuli were, Anger = 7, Disgust = 5.3, Fear = 6, Happiness = 6.9, Surprise = 7. As expected, Disgust and Fear were perceived to be the most uncanny. The removal of blendshapes affected the perceived eeriness of every emotion in a different manner. For Anger, removing the blendshapes Sneer, UpperLip, LowerLip and Puff caused the animation to be rated as more uncanny than the Basic. All of these blendshapes are an integral part of the angry expression. When these blendshapes are removed, the observer is confused when looking at the animation as the character looks angry but some of the features that we as humans, associate to anger are lacking thus making the character look strange.

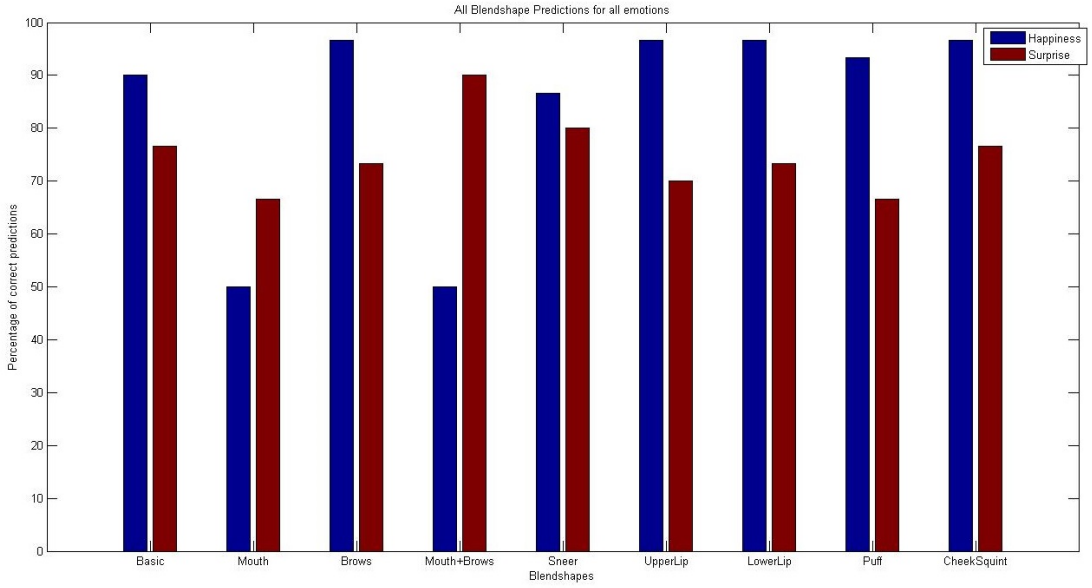

Figure 4.2: The Blendshape-wise Correct Prediction Percentages for Happiness and Surprise

This might also alter the observer's perception about the emotion. Making them question if the character is really angry or disgusted or something else entirely. This uncertainity might be the cause of the eeriness. Although, anger is a negative emotion, its familiarity rating was much higher than the other negative emotions like fear and disgust. Surprise is usually percieved as a positive emotion and in this experiment surprise was perceived as less uncanny than the others. Removing the blendshapes Brows amd CheekSquint caused the animation to perceived as more uncanny. The reason might be that, eye brows and subtle mouth movement that takes place when we gasp in surprise are one of the most important facial movements while expressing suprise. It was expected that surprise will be rated similar to fear as both comprise of similar facial movements. However, it was not rated as eerie as fear. Fear had the second lowest rating for familiarity after Disgust. As fear is considered to be the most uncanny emotion, it was expected that removing any blendshapes will make it look more uncanny. Removing any blendshape surprisingly makes the animation appear less uncanny. Although the only blendshapes that affect the perceived eeriness

significantly are UpperLip and LowerLip. As Tinwell [19] said that there is very less facial movement for fear and any mistake in the facial motion capture or animation might cause fear to be perceived as surprise. Even in this experiment, many users identified fear as surprise. Thus, the restriction of movement caused by the disablimg of the bledshapes actually made it look less strange. Disgust was rated as the most uncanny. Removing most blendshapes did not have any effect on the perceived uncanniness. The only blendshapes which caused a significant difference were UpperLip and Puff. Just like in the case of Fear the removing of these blendshapes made the animation less uncanny. The curling of the upper lip and wrinkling of the nose which are the distinguishing facial movements to express fear. These actions may may not have been captured properly or may have been captured as exagerrated. Removing the blendshapes UpperLip and Puff may have restricted the movement which was exagerrated before and made it to appear more authentic and as a result, less uncanny. Again as Tinwell [19] suggested that Disgust will continue to be uncanny regardless of the modifications made to other parts of the face. Happiness was affected the most by the removal of different blendshapes. Removing Mouth, Sneer and CheekSquint affected the percevied eeriness of the animation. The removing of these blendshapes made the animation more uncanny. The Mouth blendshape is the Smile Mouth blendshape, so removing it makes the animation look very strange, as the character is smiling, but the corners of his mouth do not move as they should. We as humans are very good at detecting false smiles and when someone is genuinely smiling, the skin below the eyes bulges and crow's feet wrinkles appear at the sides of the eyes. Removing sneer and cheek squint may cause the smile to not display some of the things that comprise a smile. Thus making it appear as a false smile, which results into the character being perceived as deceptive, with hidden intenstions and uncanny. The same theory works in the opposite way when we consider brows. Although, there are no crow's feet with the eye brow active, by removing the EyeBrow blendshape the lack of motion in the upper facial area may cause the animation to look less fake [17] [19]. Thus, decreasing the uncanniness.

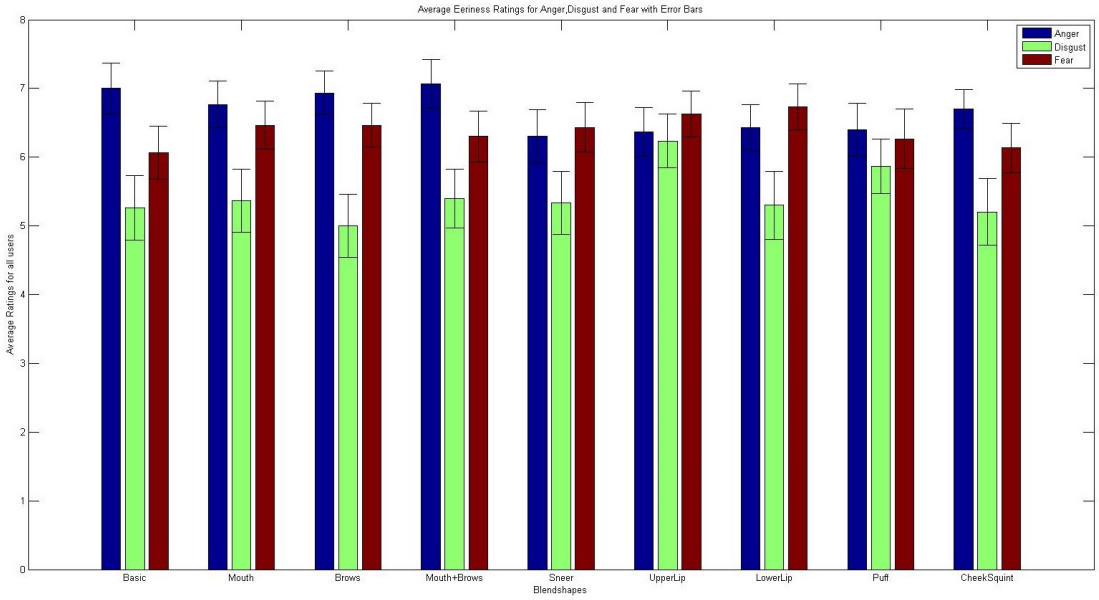

Figure 4.3: The Eeriness ratings with Error bars for Anger,Disgust and Fear

4.2.3 Human-likeness

Fig. 7 shows the average human-like ratings for Anger Disgust and Fear. Fig. 8 shows the average human-like ratings for Happiness and Surprise. The model used for this experiment, 'Macaw' was the same for all the videos. Therefore, it was expected that the ratings for the first question would not vary. However, there were differences when the blendshapes were removed and there were differences between the ratings for every emotion for the Basic stimuli(BL1). The mean human-like ratings for BL1 for Anger, Happiness and Surprise were above 7.4 out of 9. The mean human-likeness for Fear and Disgust was below 6.5. Although, the same 3D model was used the human-likeness ratings for fear and disgust were lower than the other emotions. As discussed, fear and disgust are considered more uncanny than the other emotions as these are survival emotions [19]. The removal of blendshapes make no significant difference for anger and surprise. The other emotions are affected as follows -

Fear - The removal of Mouth, Sneer, UpperLip, LowerLip, Puff and CheekSquint all causes

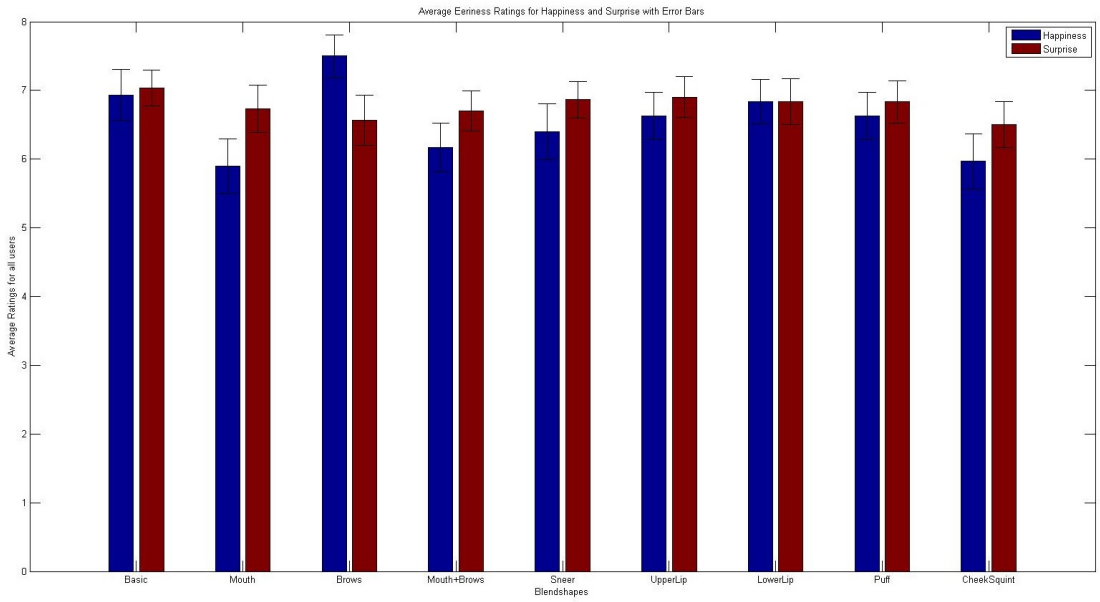

Figure 4.4: The Eeriness ratings with Error bars for Happiness and Surprise

the character to look more human-like than the Basic stimuli. The rest are insignificant. The blendshpaes which affected the perceived human-likeness control most of the lower facial region. Disgust - Only the removal of UpperLip makes a significant difference. Interestingly, curling of the upper lip is the distinguishing movement to express disgust. However removing the UpperLip blendshape makes it look more human-like instead of less human-like. The reason might be that the movement of the upper lip in the animation must be slightly off from how a human would do it. Removing the UpperLip blendshape restricted the motion of the upper lip which would have made it appear more human-like than the Basic. Again it could be because of the shortcomings of the motion capture or the acting.

Happiness - Removing the blendshapes Mouth, Sneer, Cheek Squint makes the character to look less human-like and Brows(BL3) causes it to look more human-like. The rest are insignificant. Mouth, Sneer and CheekSquint all control the lower facial region. Absence of these blendshape makes the smile look incomplete with respect to the Basic and thus make it appear less human-like. As for the Brows blendshape, the reason is similar to the uncanniess. We as humans are good at detecting a fake smile. With no crow's feet at the

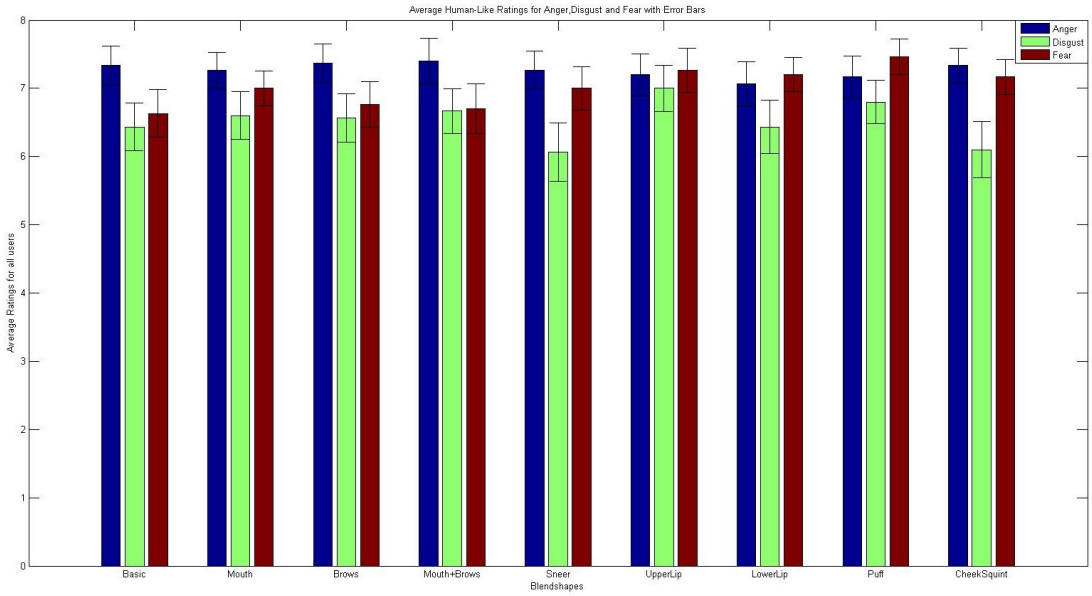

Figure 4.5: The Human-like ratings with Error bars for Anger,Disgust and Fear

side of the eyes the smile must looks fake and in this case may have appeared to be slightly off from a real human-like smile. Restricting the eye brow movement may have caused the smile to appear more human-like in comparison to the Basic video.

It was expected that there would not be stark differences between the ratings across all the videos for human-likeness. However, it was observed that the emotion and the blendshapes used change our perception of human-likeness for the same character model. Observers perceive human-likeness not only as how human-like the 3D character appears but also how human-like does it act.

4.3 Discussion

This section mentions some of the problems that I came across and some things that could have been changed.

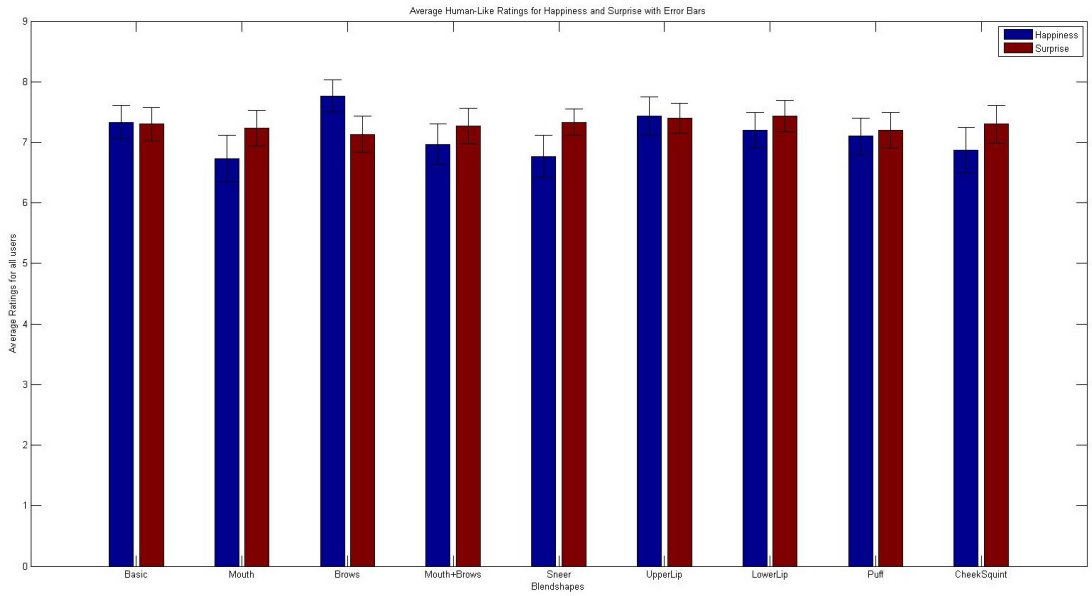

Figure 4.6: The Human-likeratings with Error bars for Happiness and Surprise

4.3.1 The Confusion with Familiarity

Throughout this thesis the words used to address the uncanny valley are eeriness, familiartiy, strangeness. However, the meaning of these words as perceived by the observers can sometimes be something other than what was intended. In general terms familiar is something that is well-known, but for this paper it is the opposite of the words weird, uncanny, eerie, bizzare, creepy or as in this case strange. This results into the participants misunderstanding the question and rating something that is well known, (in this paper, The smiling action of the avatar 'Macaw') as familiar, when they should have rated it as strange (uncanny). It is important to take into account the words used to describe the feeling that is evoked after interacting with an animated character [17]. At the same time it is also important to recognize what we do not feel towards the character. Mori, when he first coined the term *bukimi no tani* or 'the uncanny valley', he also struggled with the word which should be used to describe what is not uncanny. He used the word *Shinwakan*. This word was translated to English as 'familiar' as it was the opposite of uncanny or strange. Bartneck et al. in their paper [1] state that this may have been translated incorrectly and says

that the original concept of The Uncanny Valley may have been 'Lost in Translation' [17]. Although, the participants in this experiment were instructed about what familiarity meant, there is still a chance of misinterpretation of the word which might have affected the study. As Bartneck et al. [1] suggest the words attractiveness or likeability might be closer translation to the word *shinwakan* than the word 'familiar'.

4.3.2 Emotions

Anger, fear and disgust are considered to be the most eerie emotions, so they were included. As discussed earlier, disgust and fear are the most eerie emotions. Surprise is often perceived as fear, so it would have been interesting to see how the results for surprise compare to fear. In this experiment the results for surprise showed no similarity with fear. Surprise was perceived as more human-like and less uncanny than fear. Also, the prediction accuracy for surprise was better than fear. Happiness was included so as to have a possitive emotion and see how it compared to all the negative emotions. Sadness could have been included but seemed unnecessary. In retrospect, sadness could have been added to see the how it differred from anger, disgust and fear.

4.3.3 Blendshapes-

Some of the blendshapes that were used for this experiment could have been changed. Seeing how the removal different blendshapes affected the perceived emotion, the 4th Blendshape which was 'Mouth+Brows' did not yeild anything significant as it was a combination of Blendshape 2 and 3. More blendhspes could have been added to get better results. The cluster of blendshapes could have been formed based on the part of the face in which they lie. For example, all blendhspes which controlled the area around the lips could have been

clustered together, similarly all blendhsapes that controlled the cheek movement and for all the blendhshapes that controlled the area around the eyes and the forehead, etc. Eye movement could have been taken into consideration.
subsectionActing and Motion Capture The actor for all the videos was a student who volunteered to act for the mostion capture. Thus, the emotions expressed by the actor might not have been how the emotions actually are expressed. A professional actor would have been more suitable for this experiment and might have yeilded better results. Also, the motion capture and animation software used was glitchy at times and in some cases the intended result may not have been achieved.

To sum it up, the adding sadness, adding meaningful clusters of blendshapes and addressing the acting and motion capture would have yeilded better results.

Chapter 5

Conclusions

5.1 Current Status

The emotion expressed in the animation plays an important role in how human-like or uncanny the animation is perceived to be. As predicted, fear and disgust are the emotions which are the most uncanny and the least human-like. Removing certain blendshapes can affect the perceived human-likeness and eeriness of the animation but no blendshape can be singled out as necessary. The emotion being expressed in the animation has a significant effect on the "necessary" blendshapes. So, for disgust the sneer, upper lip and lower lip make a significant difference. That is, the region of the face that is used to express a particular emotion, decides the blendshapes that are necessary for that emotion.

Even in the case of predictions the blendshapes play an important role. But, no single blendshape is significant and the correctness of the predictions vary based on the emotion that is being expressed in the animation. Similar to human-ness and eeriness the blendshapes that affect the predictions also vary based on the emotion. The lower facial region also plays an important part in the perceived uncanniness and human-likeness and its not just the Non Verbal Communication in the upper facial region.

Some of the blendshapes, when removed, make the animation less uncanny as compared to the basic blendshape and increase the prediction accuracy but, these blendshapes vary for every emotion. For fear and disgust a lot of blendshapes made the animation look less uncanny and more human-like than the Basic. But many of these blendshapes were found to be detrimental for expressing that emotion. Ideally, removing these blendshapes should

have made the animation uncanny and decrease the prediction accuracy. This might be because of the the fact that the motion capture software was not able to track the actor accurately and/or was not able to map those facial movements on the 3D avatar accurately. This could also be because of the actor not able to enact the emotions authentically, thus making it look fake and hence uncanny. Removing these blendshapes reduced that uncanniness and made it look more authentic.

5.2 Future Work

Only seven different blendshapes were used, as one of the video stimuli had two blendshapes (Mouth and Brows) removed from it. The plan is to use other blendshapes to effectively investigate how they affect the human-ness and uncanniness and further investigate which parts of the face exactly cause the uncanny valley. In this experiment the actor just expressed one emotion at a time in the clip. It would be interesting to see how the ratings change for a complex scene involving multiple expressions and adding audio to the scene. It would also help to obeserve the changes in the ratings when the emotions change. Creating clusters of blendshapes, based on the area which they directly affect and then removing these clusters would provide more insight into the facial features which can affect the uncanny valley most significantly. Adding sadness as one of the emotions will help understad how the it affects the uncanny valley. Using different motion capture and animtaion software with more photorealistic characters will help in understanding the effect of removing blendshapes from an animation more accurately. The most important feature eyes were left out because of so much of the previous work suggesting that eyes were the most imortant feature of the face and some glitches in the motion capture software which could not determine if the eyes were open or close. This resulted into flickering eyes, including which would make the aniamtion uncanny and not human-like. If eyes are included in the blendshape clusters, it would be easier to identify the parts of the face that affect the uncanny valley.

Bibliography

[1] Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. My robotic doppelgänger-a critical look at the uncanny valley. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 269–276. IEEE, 2009.

[2] David Bennett. The faces of "the polar express". In *ACM SIGGRAPH 2005 Courses*, SIGGRAPH '05, New York, NY, USA, 2005. ACM.

[3] Erika Chuang and Chris Bregler. Performance driven facial animation using blendshape interpolation. *Computer Science Technical Report, Stanford University*, 2(2):3, 2002.

[4] Vanderson Dill, Laura Mattos Flach, Rafael Hocevar, Christian Lykawka, Soraia R. Musse, and Márcio Sarroglia Pinho. *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings*, chapter Evaluation of the Uncanny Valley in CG Characters, pages 511–513. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[5] JM Donovan. Review of facial attractiveness: Evolutionary, cognitive, and social perspectives edited by gillian rhodes and leslie a. zebrowitz. *Human Nature Review*, 3:226–228, 2003.

[6] S. Freud and J. Trachey. The uncanny. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, volume 17, pages 217–256. Hogarth Post London, 1919.

[7] Robert D Green, Karl F MacDorman, Chin-Chang Ho, and Sandosh Vasudevan. Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior*, 24(5):2456–2474, 2008.

[8] Jessica Hodgins, Sophie Jörg, Carol O'Sullivan, Sang Il Park, and Moshe Mahler. The saliency of anomalies in animated human characters. *ACM Trans. Appl. Percept.*, 7(4):22:1–22:14, July 2010.

[9] Karl F MacDorman. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*, pages 26–29, 2006.

[10] Karl F. MacDorman, Robert D. Green, Chin-Chang Ho, and Clinton T. Koch. Too real for comfort? uncanny responses to computer generated faces. *Comput. Hum. Behav.*, 25(3):695–710, May 2009.

[11] Karl F MacDorman and Hiroshi Ishiguro. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3):297–337, 2006.

[12] Takashi Minato, Michihiro Shimada, Hiroshi Ishiguro, and Shoji Itakura. Development of an android robot for studying human-robot interaction. In *Innovations in applied artificial intelligence*, pages 424–434. Springer Berlin Heidelberg, 2004.

[13] Masahiro Mori, Karl F. MacDorman (Translator), and Takashi Minato (Translator). The uncanny valley. *Energy*, 7(4):33–35, 2005.

[14] Jun'ichiro Seyama and Ruth S. Nagayama. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoper. Virtual Environ.*, 16(4):337–351, August 2007.

[15] William B Swarm Jr. Self-verification: Bringing social reality into harmony with the self. *Social psychological perspectives on the self*, 2:33–66, 1983.

[16] A. Tinwell, M. Grimshaw, and D. Abdel-Nabi. The uncanny valley and nonverbal communication in virtual characters. In *Nonverbal Communication In Virtual Worlds*, pages 325–342. ETC Press, 2014.

[17] Angela Tinwell. The uncanny valley in games and animation, 2014.

[18] Angela Tinwell and Mark Grimshaw. Bridging the uncanny: An impossible traverse? In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, pages 66–73. ACM, 2009.

[19] Angela Tinwell, Mark Grimshaw, Debbie Abdel Nabi, and Andrew Williams. Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior*, 27(2):741–749, 2011.

[20] Angela Tinwell, Mark Grimshaw, and Andrew Williams. Uncanny behaviour in survival horror games. *Journal of Gaming &# 38; Virtual Worlds*, 2(1):3–25, 2010.

[21] Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. Realtime performance-based facial animation. In *ACM Transactions on Graphics (TOG)*, volume 30, page 77. ACM, 2011.

[22] L. Winerman. "the minds mirror". In *Monitor on Psychology*, volume 36, pages 48–50. 2005.

Appendix A

All Graphs from Analysis

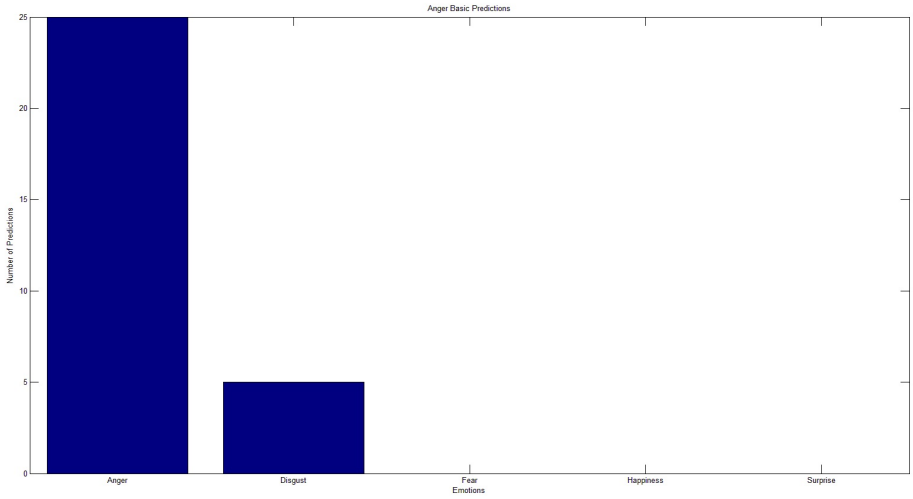

Figure A.1: How the Basic video for Anger was Predicted

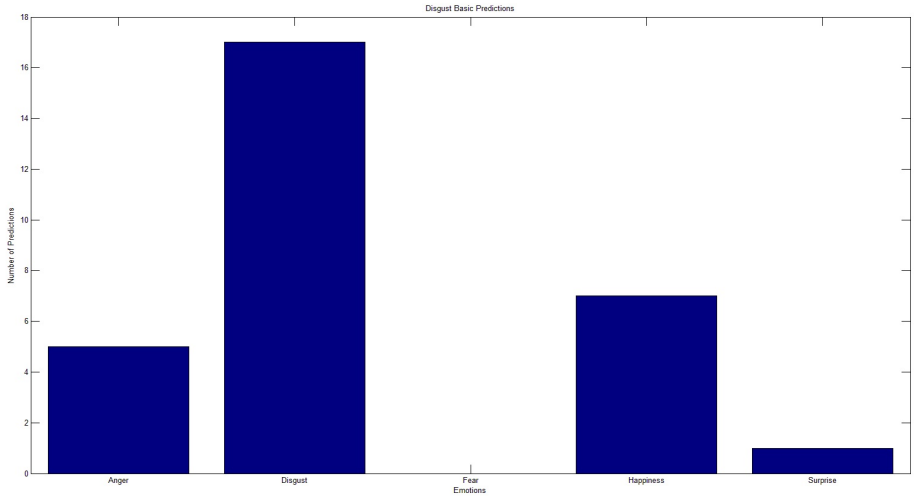

Figure A.2: How the Basic video for Disgust was Predicted

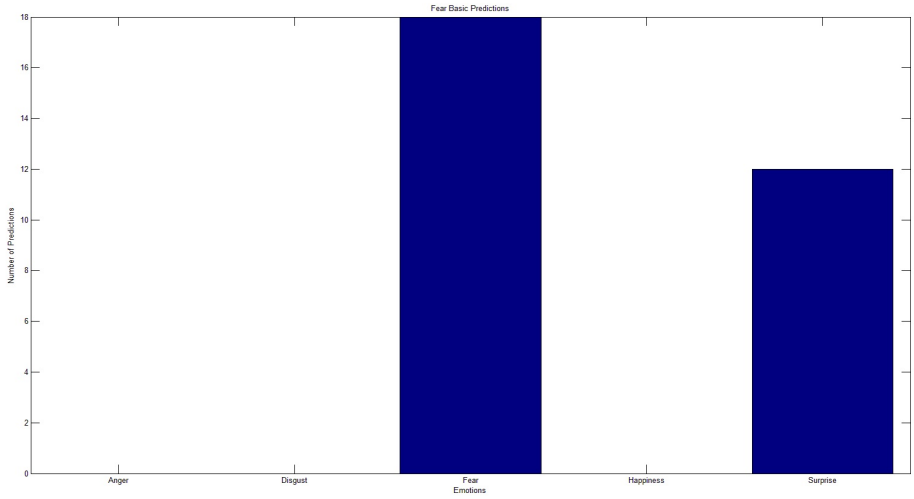

Figure A.3: How the Basic video for Fear was Predicted

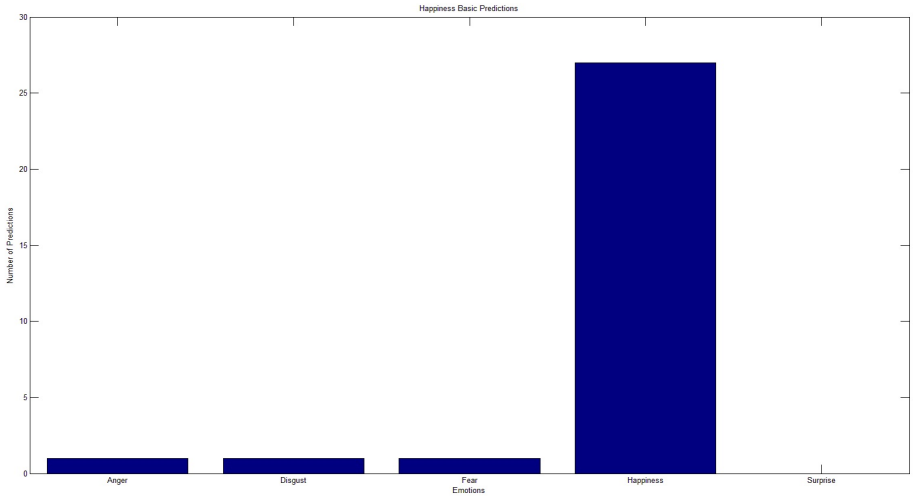

Figure A.4: How the Basic video for Happiness was Predicted

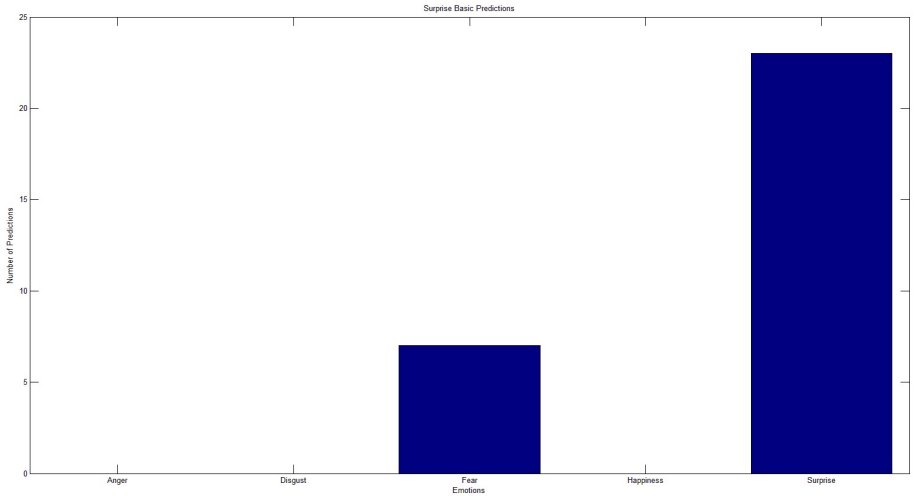

Figure A.5: How the Basic video for Surprise was Predicted

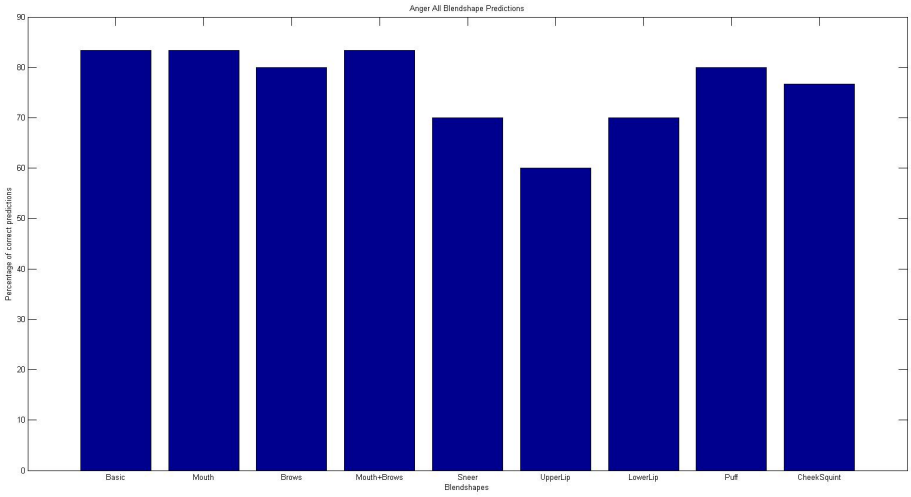

Figure A.6: Blendshape wise correct predictions for Anger

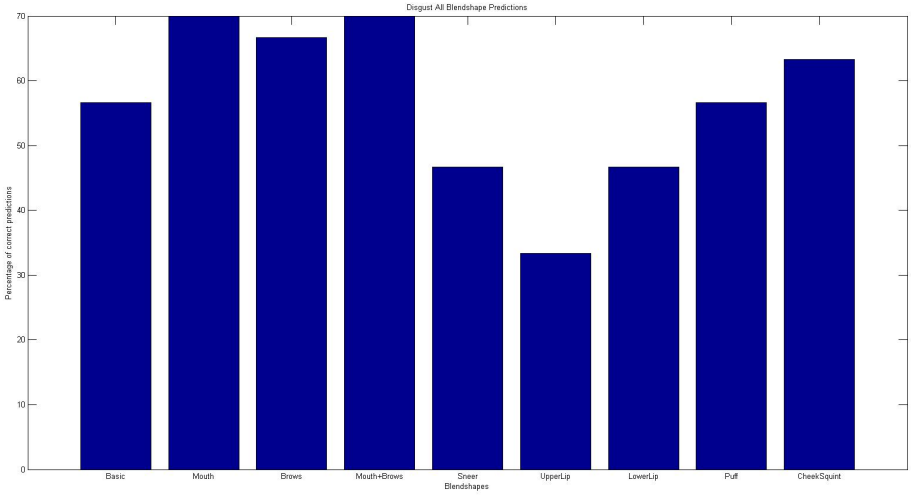

Figure A.7: Blendshape wise correct predictions for Disgust

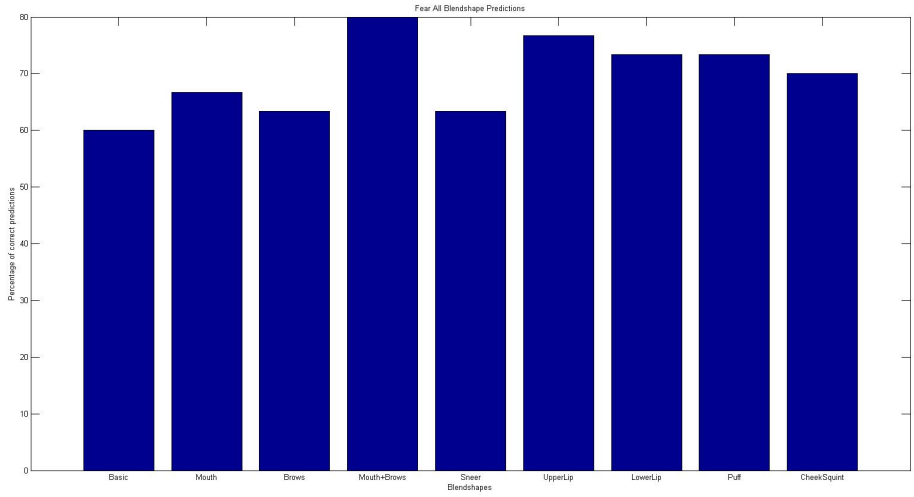

Figure A.8: Blendshape wise correct predictions for Fear

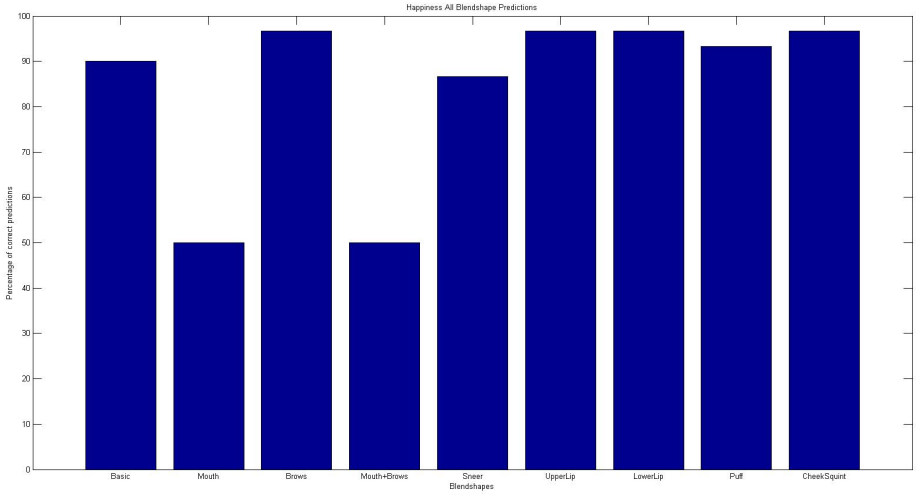

Figure A.9: Blendshape wise correct predictions for Happiness

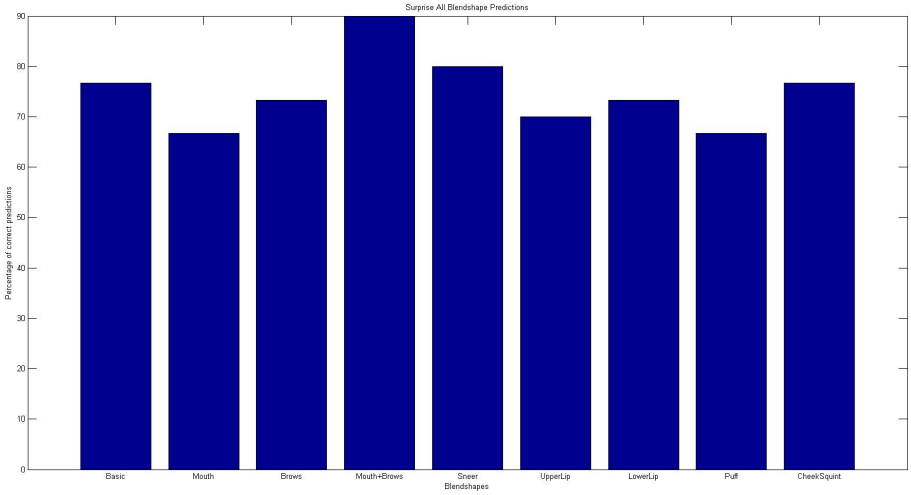

Figure A.10: Blendshape wise correct predictions for Surprise

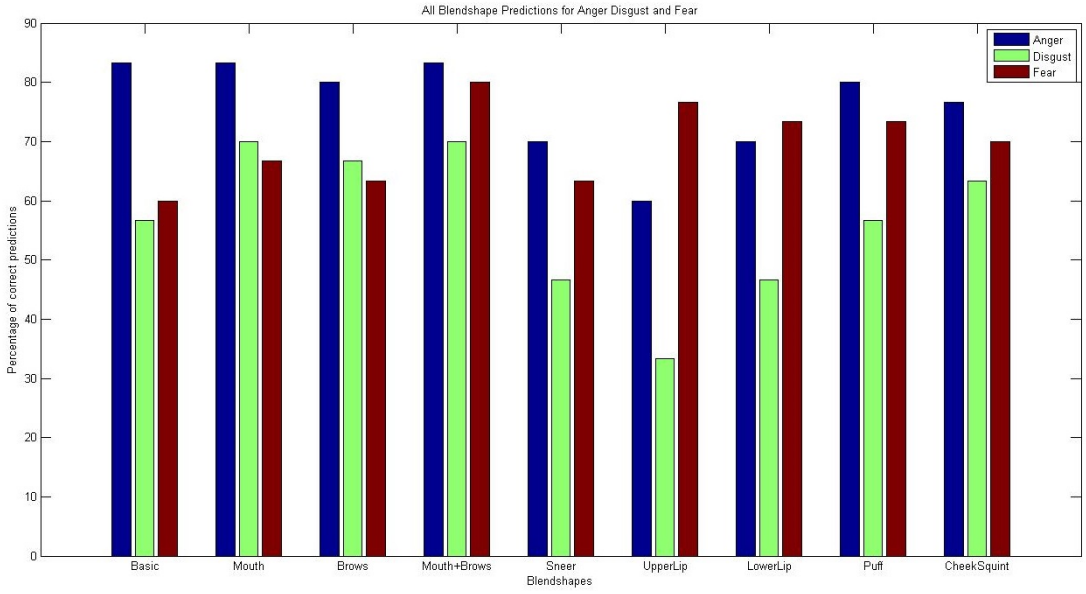

Figure A.11: Blendshape wise correct predictions for Anger, Disgust and Fear

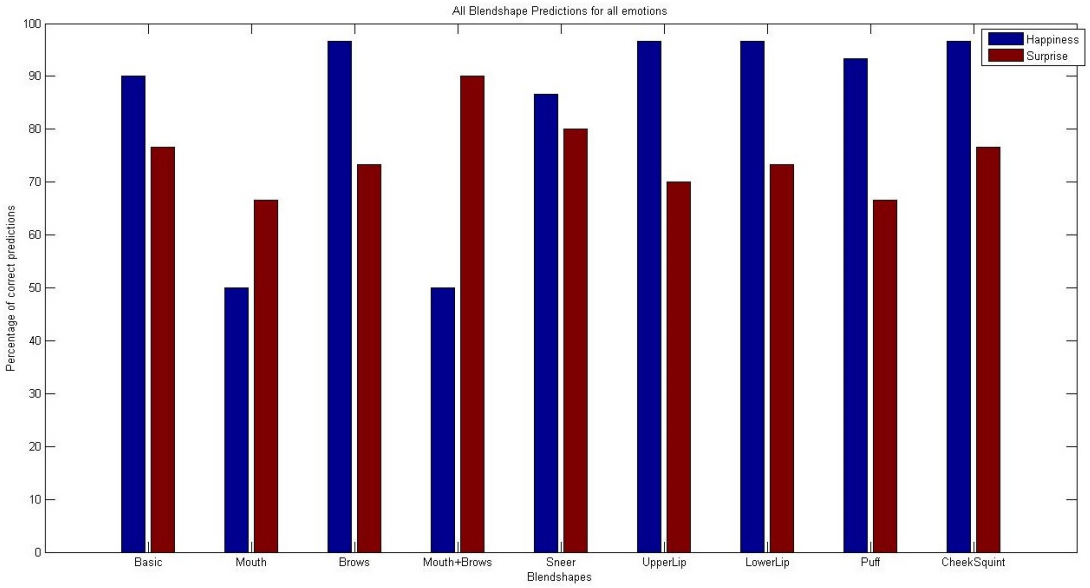

Figure A.12: Blendshape wise correct predictions for Happiness and Surprise

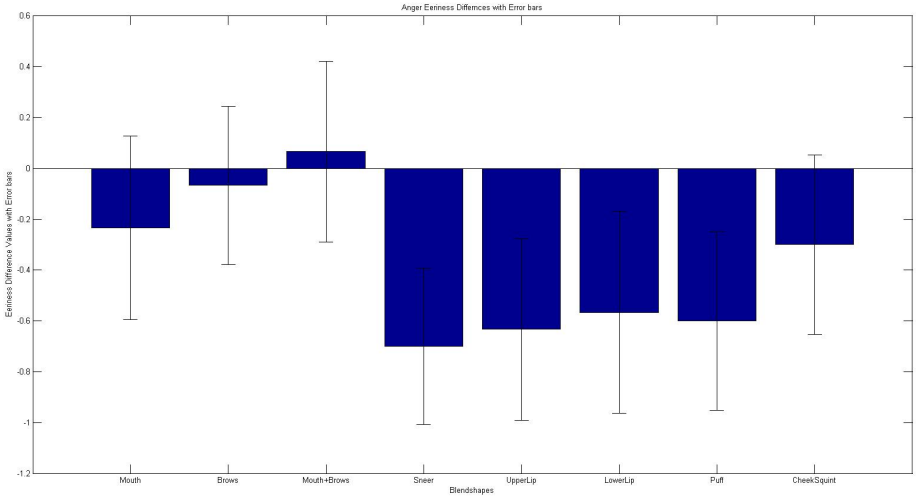

Figure A.13: Eeriness differences with error bars for Anger

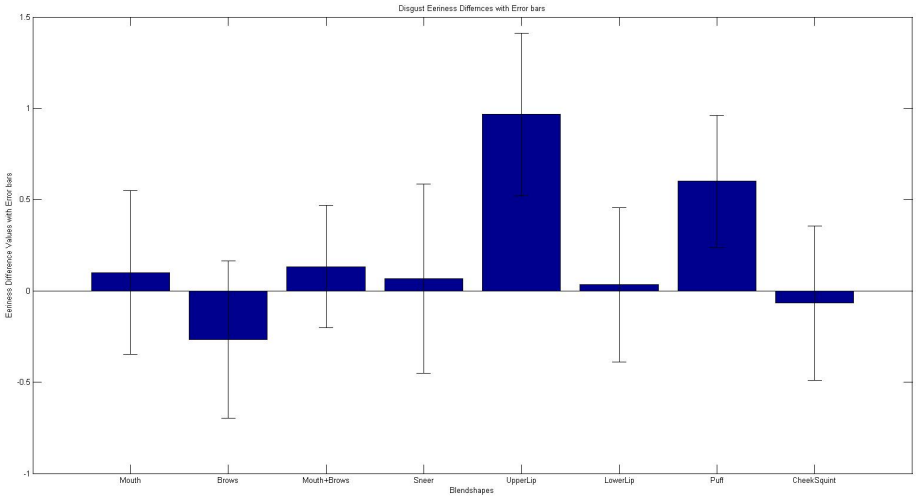

Figure A.14: Eeriness differences with error bars for Disgust

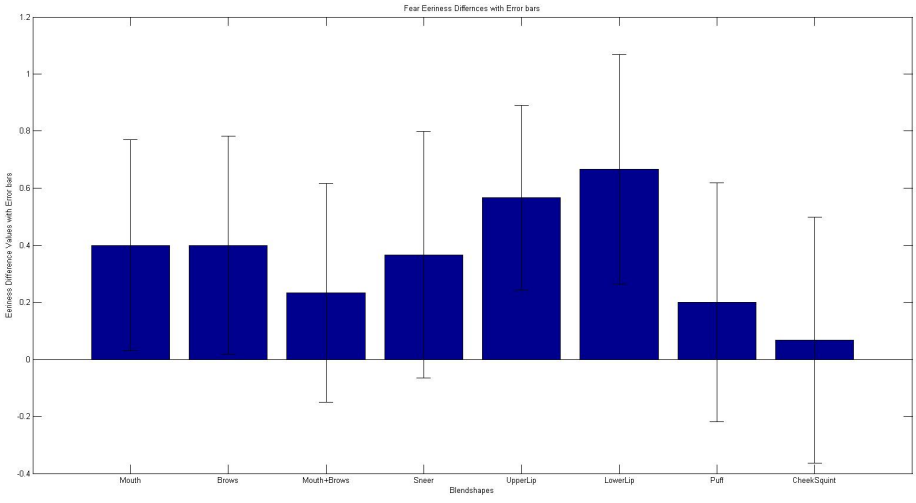

Figure A.15: Eeriness differences with error bars for Fear

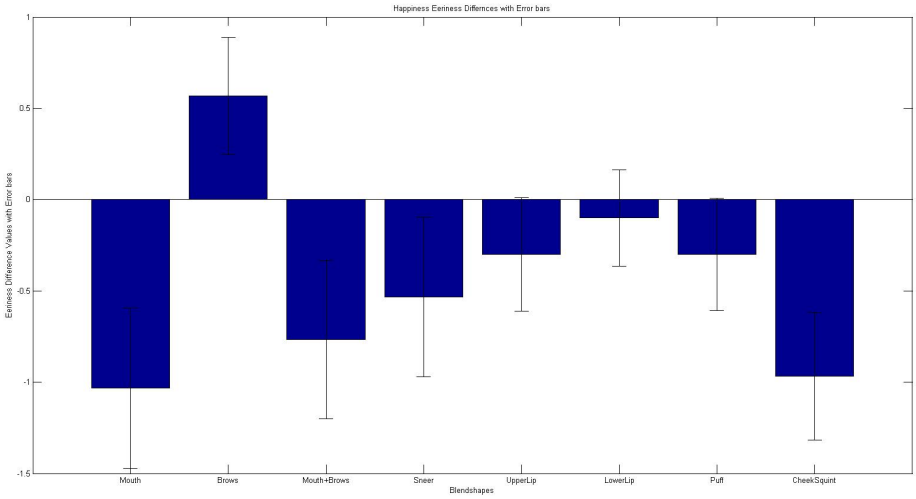

Figure A.16: Eeriness differences with error bars for Happiness

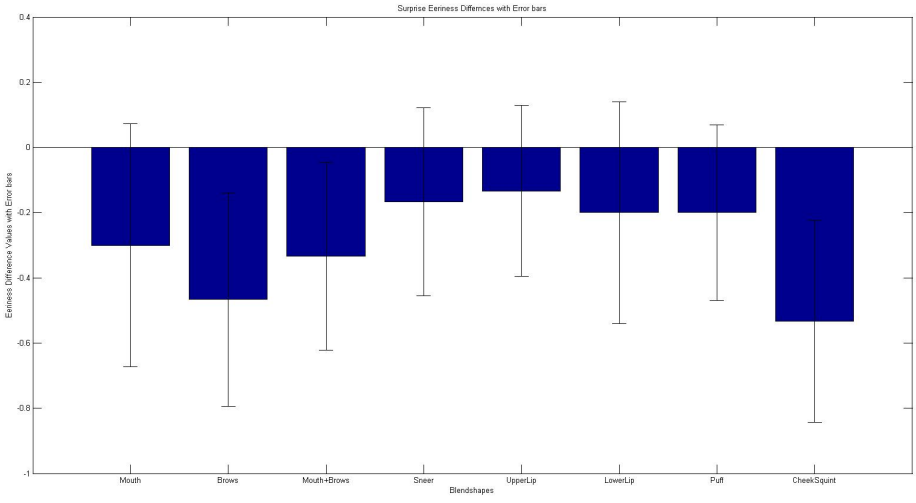

Figure A.17: Eeriness differences with error bars for Surprise

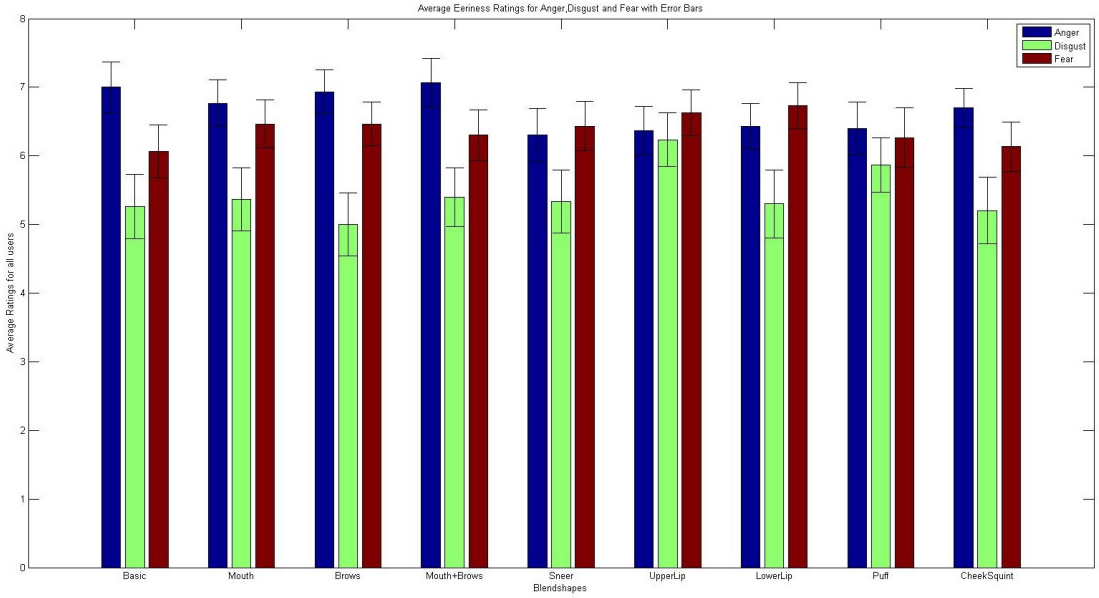

Figure A.18: Eeriness average ratings with error bars for Anger, Digust and Fear

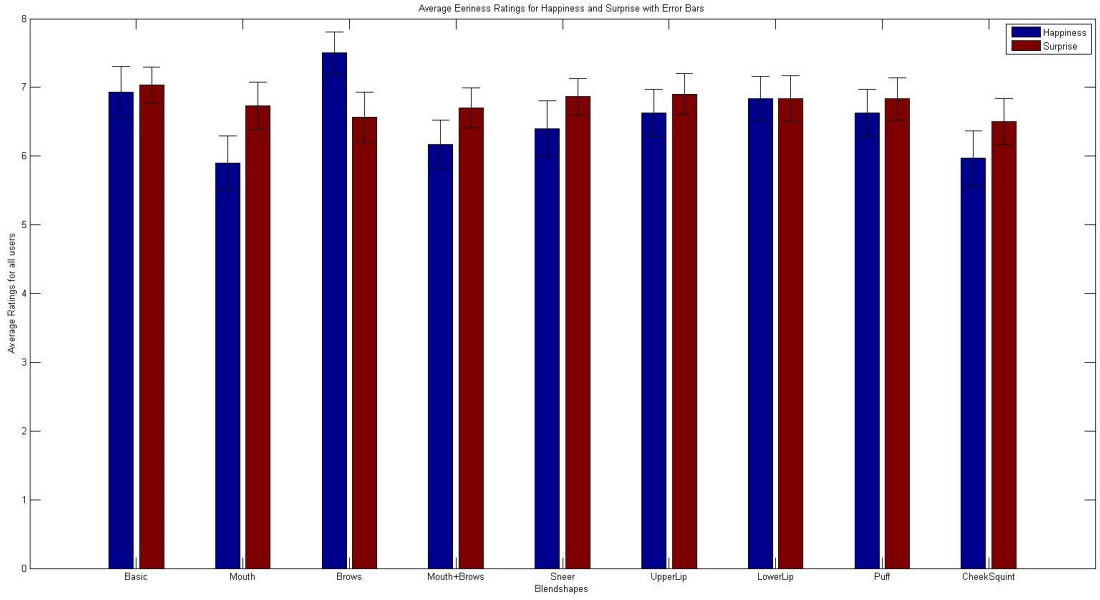

Figure A.19: Eeriness average ratings with error bars for Happiness and Surprise

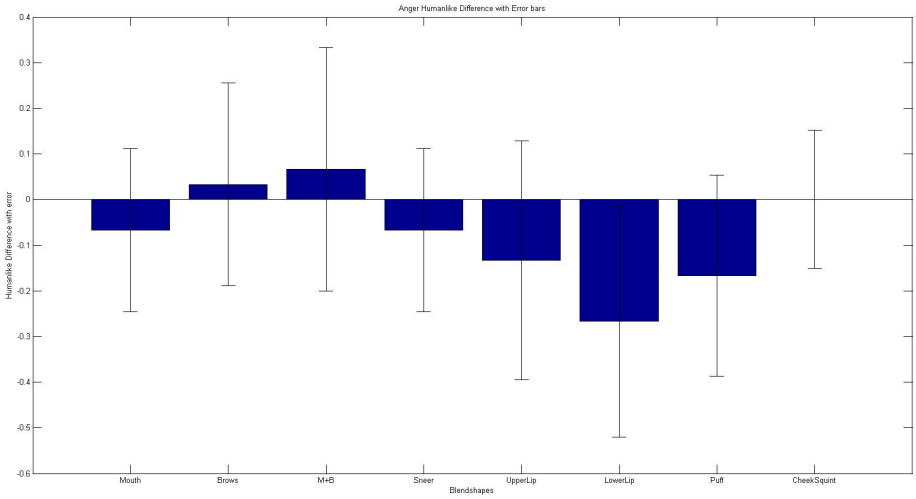

Figure A.20: Human-like differences with error bars for Anger

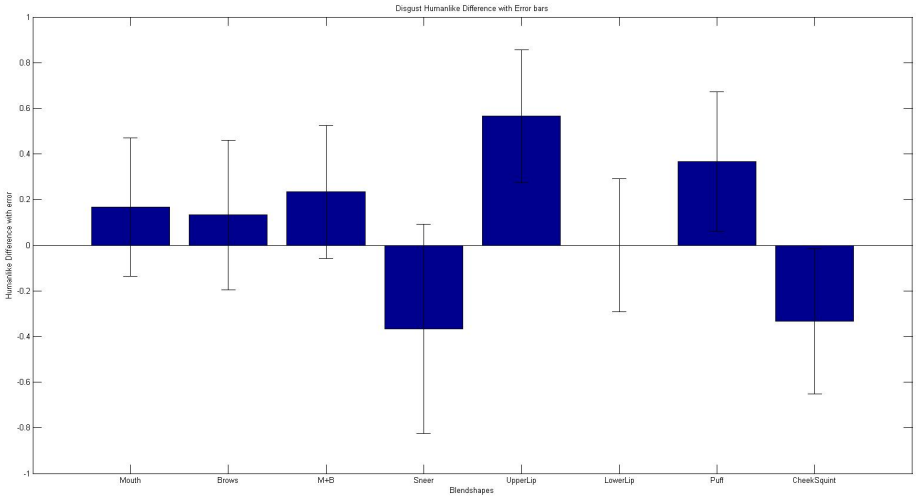

Figure A.21: Human-like differences with error bars for Disgust

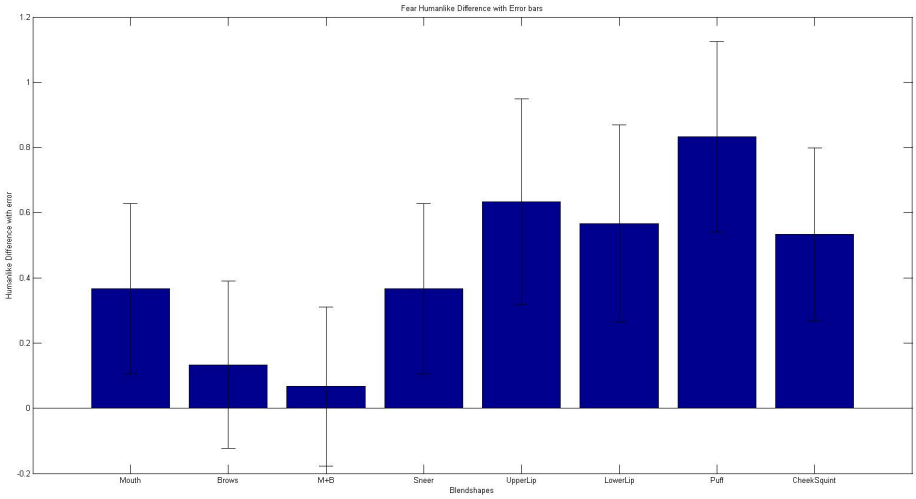

Figure A.22: Human-like differences with error bars for Fear

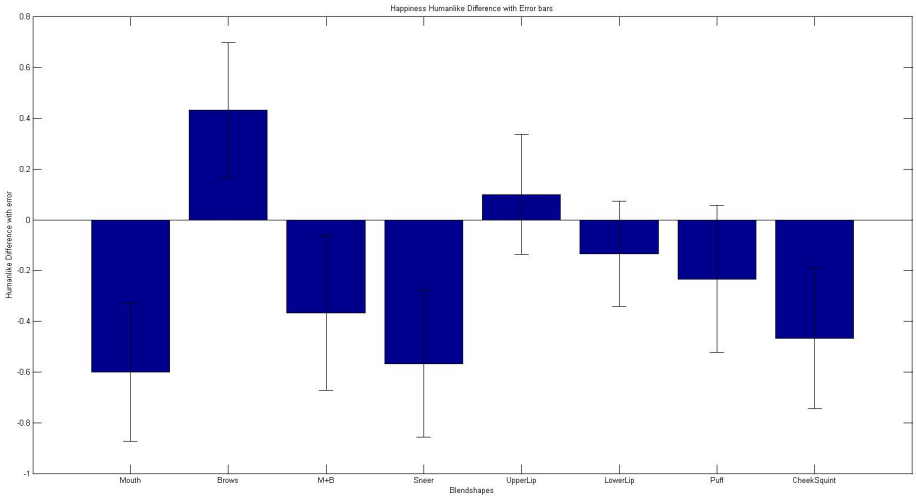

Figure A.23: Human-like differences with error bars for Happiness

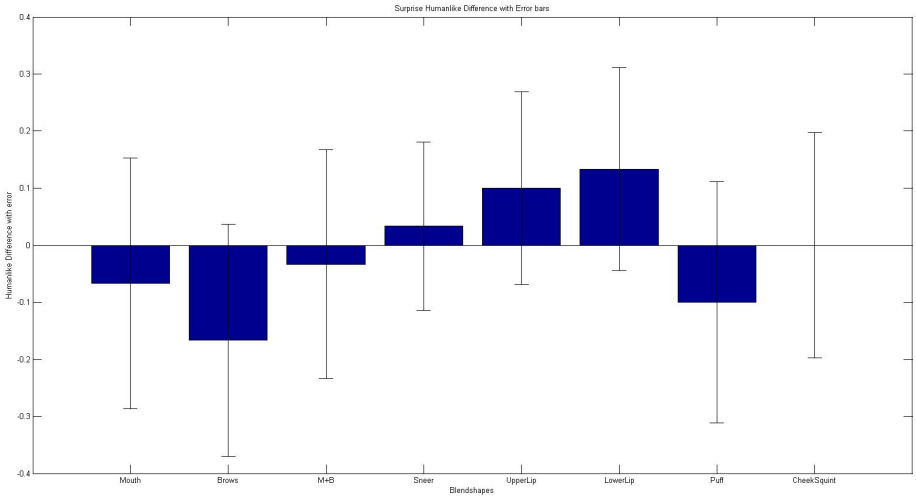

Figure A.24: Human-like differences with error bars for Surprise

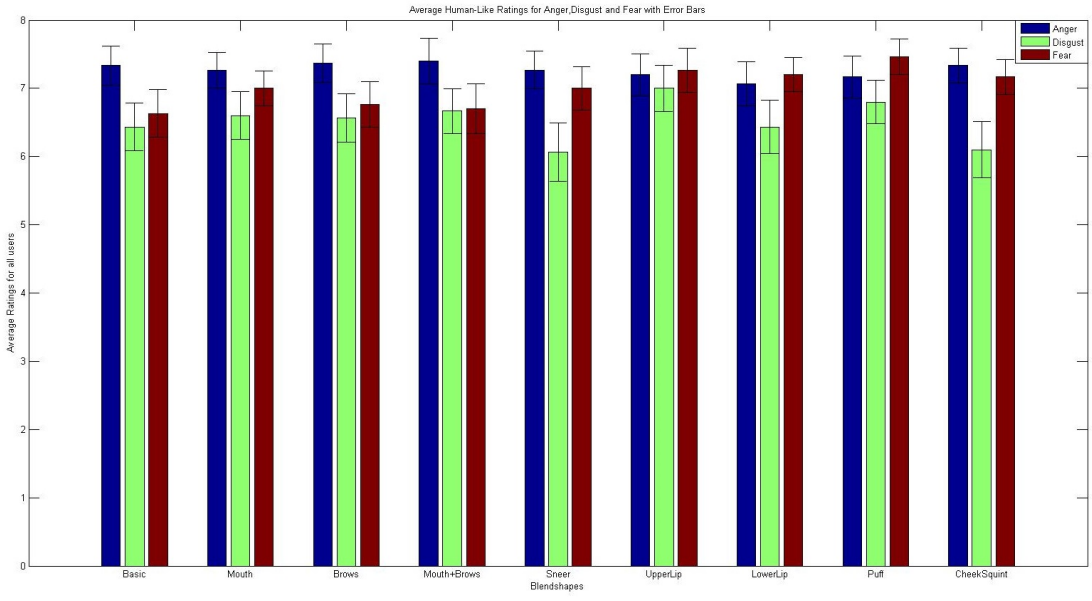

Figure A.25: Human-like ratings with error bars for Anger, Digust and Fear

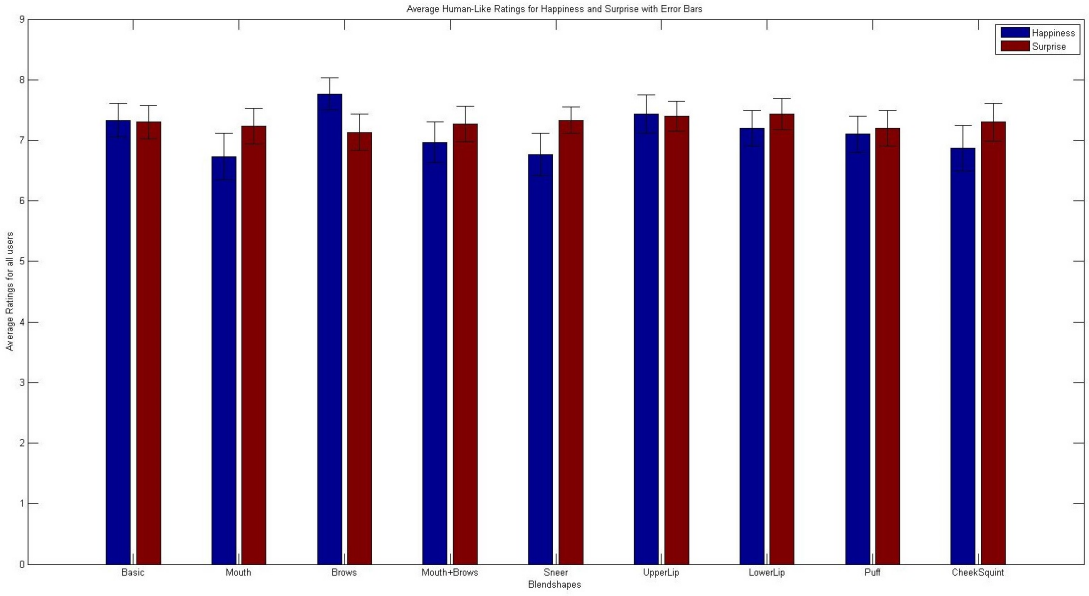

Figure A.26: Human-like ratings with error bars for Happiness and Surprise