Investigating Interactions Between Machines: A Case Study Using Facial Expression Recognition and Virtual Avatars

By

Keenan Halliday van Rooyen
1386104

Master of Arts by Research in the field of Digital Arts

Supervised by: Dr. Tegan Bristow, with assistance by Dr. Pravesh Ranchod

February 2022

Ethics Application Number: HRECNM21-11-008

Ethics Waiver Letter Number: HRECNMW21/11/10

A dissertation submitted to the Wits School of Arts, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Arts

Abstract

Computer vision is a field of artificial intelligence which revolves around enabling machines to derive meaningful information from visual inputs (Zafeiriou, Zhang, and Zhang, 2015: 1). Researchers have shown a focused interest on using computer vision to develop a machine which can both interpret and classify human behaviour, movements, and emotions through only visual information (Zafeiriou et al., 2015: 2). What was found through a literature review was that there is a gap in knowledge in computer vision for studies which do not rely heavily on human analysis. The presented research study aimed to work towards filling this gap by investigating an interaction between the goals of computer vision through using virtual avatars from the *Animaze* library to imitate human emotional expressions that were then analysed by *FaceReader's* expression recognition component. This study created a machine-based interaction which analysed how changes of the facial features on a virtual avatar could alter the analysis of a computer vision program interpreting emotional expressions in the face. The results of this experiment were separated into three sections to answer three guiding research questions and to provide a wider scope of analysis. An analysis of FaceReader's results presented several findings surrounding computer vision and virtual avatar interactions, with the most notable finding being that even slight changes in the shape and size of facial features on a virtual avatar can produce vastly different emotional expression readings. It was also found that it is possible to influence *FaceReader's* expression analysis to produce higher, or lower, intensity averages of the emotional expressions through the use of specific facial features. Overall, the presented research found information that one could argue was already proven for human facial recognition, but value was provided in this research being potentially used as a steppingstone towards further development in the use of facial recognition on virtual avatars.

Keywords: Computer Vision; Facial Recognition; Virtual Avatar; Emotional Expressions; FaceReader; Animaze

<u>Declaration</u>

I, KEENAN HALLIDAY VAN ROOYEN, hereby declare the contents of this research dissertation to be my own work. This dissertation is submitted for the degree of Master of Arts in Digital Arts at the University of the Witwatersrand. This work has not been submitted to any other university, or for any other degree.

____________________ <u>14 February 2022</u>

Keenan Halliday van Rooyen

Contents

List of Figures

List of Tables & Graphs

Glossary of Terms

Algorithm: set of rules to be followed by a machine when doing calculations or problem-solving, usually during machine learning related tasks (Magdin and Prikler, 2017: 2).

Artificial Intelligence: The development of machines and their ability to perform tasks commonly associated with intelligent or sentient beings (Pasquale, Percannella, & Vento, 2014).

Computer Vision: the use of artificial intelligence to enable machines to identify and gather meaningful information from visual inputs (Zafeiriou, Zhang, and Zhang, 2015: 1).

Dataset: A collection of data that shares a common attribute which is used in training models (Magdin, Benko, and Koprada, 2019: 4).

Deep Learning: An artificial intelligence technique that teaches a computer to filter inputs through layers to learn how to predict and classify information (Kuo et al., 2018).

Facial Expression Recognition: Using computer vision to recognise the emotional expressions being shown on a human face (Li and Deng, 2018).

Facial Recognition: A way of recognising the human face through computer vision and technology (Li and Deng, 2018).

Machine: A computer or similar technological device (Zafeiriou, Zhang, and Zhang, 2015: 1).

Machine Learning: The use of Artificial Intelligence to provide a machine the ability to learn and improve by itself using experiences (Pasquale, Percannella, & Vento, 2014).

Model: A computer program or file that has been trained to recognize certain types of patterns through machine learning (Rincon, Costa, Carrascosa, Novais, and Julian, 2019).

Neural Network: A collection or series of algorithms that are connected in a way to mimic the human brain to adapt and learn by itself (Magdin et al., 2019: 4).

Training: The teaching of an algorithm or model to identify specific data or attributes (Magdin et al., 2019: 4).

SECTION ONE: INTRODUCTION

1.1 Introduction

Computer vision is a field of artificial intelligence which revolves around enabling machines to derive meaningful information from visual inputs (Zafeiriou, Zhang, and Zhang, 2015: 1). Using computer vision, researchers have shown great interest in creating a machine which can interpret and classify human behaviour, movements, and emotions through visual information (Zafeiriou et al., 2015: 2). While there are numerous studies which focus on other aspects of human features, such as: posture, limb movement, heart rates, the analysis via computer vision of facial features has been shown to be invaluable when researching emotion identification and expression classification (Song, Shi, & Reed, 2020; Magdin, Turcani, & Hudec, 2016)

Being able to interpret how other people feel is a vital element of social interactions as understanding the emotional state of others allows humans to communicate more effectively (Magdin et al., 2016: 62). Since the publication of Charles Darwin's book, *The Expression of the Emotions in Man and Animals* (Darwin, 1872), the face has been considered to play an integral role in non-verbal communication. While verbal communication is important, visual cues assist humans when interpreting each other's emotional states. In his book, Darwin proposed that facial movements are not necessarily caused by emotional states but rather that certain expressions became associated with emotions as a by-product of survival mechanisms (Parkinson, 2005: 20). This meant that facial expressions were believed to not have a direct link to emotions, rather it was believed that humans adapted to portraying facial expressions to communicate more effectively. While Darwin believed there was no direct link between facial expressions and emotions, his theories on facial expressions became fundamental in Tomkins' "Basic Emotion Theory" (BET) (Tomkins, 1962; Dupré, Krumhuber, Küster, & McKeown, 2019: 3).

BET assumes that there are signature facial expressions which portray a limited number of "basic emotions" (known as: happiness, sadness, anger, fear, surprise, and disgust) (Dupré et al., 2019: 8) in a universally understood way. This theory is rooted in the belief that facial expressions were adapted to be functional and were necessary for human survival, making them instinctually understandable by humans even across different cultures (Ekman, 1970: 152). BET's validity has since been questioned in research studies conducted in the last few years (Barrett & Wager, 2006; Di Fiore, Quax, Vanaken, Lamotte, & Van Reeth, 2014) which provide evidence that different cultures have varying norms for non-verbal communication which affect how the six basic emotions are interpreted and portrayed. Even though there has been questioning of his studies, Tomkins' BET theory remains apparent in the development of computer vision and developing machines to detect human emotion (Dupré et al., 2019: 8).

1.2 Research Background

The methods surrounding emotion classification have found a common approach which enable machines to analyse facial feature movements and group them into facial expressions. This approach consists of three steps: Identification, Detection, and Classification (Loijens, & Krips, 2021: 3). The first step involves identifying and tracking a human face from the input (image, video) based on how the machine was designed to recognise faces. The second step involves detecting landmarks on the face and the positioning of these landmarks are tracked. These landmarks are used to outline important facial features, such as the mouth and eyes. The final step is where the machine will use the landmarks to classify the facial expression being portrayed by matching the detected landmark patterns to the patterns the machine was trained to recognise (Deshmukh, Paygude, & Jagtap, 2017: 220). The final step in the facial expression recognition process is where one is able to see the influence BET has in the use of computers to detect human emotions.

The classification of emotions through computer vision has commonly been limited to the six basic emotions of BET (Lozano-Monasor, López Bonal, Fernández-Caballero & Vigo-Bustos, 2014: 1). The use of only these emotions is due to the assumption that emotional expressions correspond to specific patterns of facial activity (Magdin & Prikler, 2017: 2). This method has been proven to be effective in assisting machines when interpreting human emotions through facial features and studies using this method have found that through a person's facial features and expressions, it is possible for a computer program to accurately estimate their emotional state when focusing on the six basic emotions (Lozano-Monasor et al. 2014: 1). Some studies further tested this and found that it is possible to interpret emotional information through isolating the positioning of individual facial features (e.g., mouth, eyes, eyebrows) and analysing them, rather than their entire face as a whole.

Through the understanding that facial features can be used when interpreting human emotional states, there came an interest in assisting machines to not only examine these human emotions but to have them react and portray these emotions as well. Several studies have investigated this concept through the use of computer-generated graphical representations of a human, referred to as virtual avatars (Bernal and Maes, 2017; Di Fiore et al., 2014). These studies would be performed through either motion capture programs, where the virtual avatar would use real-time tracking to mimic a human counterpart, or through modelling software, which would allow the researchers to change the positioning of key facial features to specific locations to mimic human facial expressions. Testing would be conducted through human participants classifying the virtual avatars according to what emotional state they believed the avatar was portraying (Faita, Vanni, Lorenzini, Carrozzino, Tanca and Bergamasco, 2015: 2). This field of research has shown a noteworthy ability within the virtual avatars to simulate human emotions through the BET understanding of how human facial features display emotions (Faita et al., 2015: 3). A point of interest that these studies produced was that there were concerns in how dynamic or natural the emotional expressions were in the virtual avatars as human participants were found to recognise posed emotions in virtual avatars and classify them as one of the 6 basic emotions.

Computer vision has advanced to be useable through still images, real-time video feeds, and video recordings, allowing the technology to be used in several facets of daily life (e.g., education, medicine, security), and computer graphics have developed exponentially with motion capture software, publicly available graphics processors, and powerful everyday devices enabling the growth of computer graphics (Geraets, Tuente, Lestestuiver, van Beilen, Nijman, Marsman, & Veling, 2021: 1). Although, even with these advancements the majority of the studies around computer vision and facial expressions through graphical representations were found to be disconnected. Facial expression recognition programs are tested on human participants (Lozano-Monasor et al. 2014), and human input has been required for engaging with and analysing how well machines can portray emotions through virtual avatars (Di Fiore et al. 2014). There are few noteworthy studies conducted which address how well a machine is able to interact with and interpret facial expressions produced by another machine.

The presented research aims to address this concern with an investigation that will analyse the interaction between machines as computer vision is used to analyse the facial expressions of virtual avatars. The intent of this research is **not** to investigate how well computer vision analyses the emotions displayed by a virtual avatar, rather this research will focus on analysing how the facial features of a virtual avatar would effect emotion expression recognition. Virtual avatars are able to simulate any human facial feature and combine them with a wide range of motions and movements that a real human would not be able to imitate (Di Fiore et al. 2014: 4). With this wide range of facial feature choices and movements, there is also the possibility that virtual avatars are able to influence the recognition of emotions through influencing the personal bias. Personal bias is when certain facial features would tend to make people feel a set of emotions are being expressed even when they are not present (Loijens and Krips, 2021: 5). Tilted down eyebrows would be bias towards portraying anger, drooping mouths would be seen as sadness, examples such as these are possible in virtual avatars, and they could possibly be pushed further to create a greater bias. This reveals the potential within the virtual avatars to be modelled with facial features that have a personal bias toward, or away from, certain expressions being recognised.

In popular media and computer graphics we can see that character's shy away from looking too realistic and tend to have larger features to portray their emotional expressions through exaggeration. Abnormally large eyes are a common example of this exaggeration, as humans gain a lot of emotional information through information gathered in the eyes and their movements. Many virtual characters are also able to portray emotional expressions clearly without a detailed nose or simple dots for eyes, meaning exaggeration through minimalism is also possible (Bartneck, 2002). There are countless options of different combinations of facial features that can be applied to a virtual avatar, and their effectiveness in portraying emotional expressions is a focal point of this research project. The facial features used by the virtual avatar for this research study were found in an extensive virtual avatar library known as *Animaze* (2021). *Animaze* is a commercially available program which uses motion capture technology to control the movements of a virtual avatar's facial features while offering a wide selection of avatar designs and facial features. The *Animaze* virtual avatars will be analysed using *FaceReader*, a facial analysis program which uses facial recognition software to identify emotional expressions through detection of facial features. *Animaze* and *FaceReader* are advanced programs, and their full details and description of use are provided in **Section 3.**

1.3 Research Question and Aims

1.3.1 Research question

"How would variations in the facial features of virtual avatars affect facial expression recognition: a study using the Animaze *Library of virtual avatars with* Face Reader's *facial expression recognition component."*

1.3.2 Research Significance

At the time that of this thesis being written, there were few noteworthy studies that investigated the interaction of computer vision with virtual avatars. Studies were limited to either being solely computer vision based, where a machine evaluated and classified expressions on a real human face (Magdin, Benko, and Koprada, 2019), or the studies were done through humans interpreting the expressions being shown on a virtual avatar (Bernal and Maes, 2017). This gap in information is becoming problematic as we move further into the modern age where the use of virtual avatars may become a primary form of commination as we transition into a world similar to the recently announced *MetaVerse* (Lei and Ratan, 2021). Virtual avatars are complex and have become more and more realistic, but this realism is from a human perspective. When the human input is removed, there is little information hinting towards what a machine will do when tasked with interpreting a virtual avatar's emotional expressions rather than the human face it was trained on. The presented research in this study aims to investigate a small section of this interaction of machines, by using computer vision to interpret the emotional expressions created by various virtual avatars. Through this investigation, this research intends to build on the knowledge base for virtual avatars with different designs expressing emotions and how these emotional expressions are interpreted by another machine. The virtual avatars will be customised and presented using *Animaze* (2021), and the emotional expression recognition will be conducted through *FaceReader* (2022). Both of these materials are discussed in depth in **Section 3** and **Section 4.**

1.3.3 Research Aims

This research project was designed around investigating how the facial expressions detected through *FaceReader* would be affected by changing the facial features of a virtual avatar in *Animaze*. The intent is to use this research project to answer the following research questions:

1. How would changes in the shapes and sizes of key facial features (nose, jaw, eyes, eyebrows, ears, mouth) affect facial expression recognition?
2. Which facial features would be the focal point for *FaceReader* when identifying specific emotional expressions and their intensities?
3. Is it possible to create an avatar in *Animaze* that causes *FaceReader* to produce higher, or lower, emotional expression intensities? If so, how would this be done?"

This meant that previous studies were focused on either analysing avatars or using machines to analyse humans, not both. Secondly, I emphasise that throughout this investigation that no human judges were used for classifying the emotional expressions or to evaluate the accuracy of either the expressions evaluated by *FaceReader*, or the expressions portrayed by *Animaze*. The intent was not to investigate expression accuracy, rather this investigation aimed to identify how variations in facial features would affect expression classification. As explained in Section 3, the same virtual avatar is analysed throughout this study but there are slight variations in the facial features used. The comparisons are done between the same avatar with slight variations, meaning the analysis will not be on how accurately the virtual avatar was able to imitate the emotional expressions of a human counterpart but rather a comparison of how FaceReader's expression analysis will change with the facial feature changes.

1.4 Section overview

Section One of this study presented an introduction to the work that will be presented throughout. This was done through an outline of the research background and motivations which lead to the chosen research questions. The significance of the presented research was also discussed followed by an outline of the research aims presented for this study.

Section Two presents the theoretical foundation for this study which begins with a discussion on computer vision and facial expression recognition. The history and methods surrounding these concepts are discussed in-depth, followed by a discussion on the challenges in research surrounding computer vision. This section continues by discussing the history and research surrounding deep learning and artificial intelligence, and what challenges were found in this field of knowledge.

Section Three focuses on discussing the focal points of this study on computer vision and virtual avatars through presenting research on facial animation and expression recognition. The various approaches to facial animation are presented in this section, followed by the approaches to computer vision and facial recognition programs. This section concludes with a summary of findings surrounding emotional expressions in the face to help better understand the approaches to expression recognition in this study and how this study will be interpreting the virtual avatars.

Section Four presents the research methodology of this study by referencing the methods and materials used to conduct the presented research. The materials section examines the different programs used for this study, with motivation as to why these programs were used, and the exact system specifications of the machine used to conduct this study to better enable recreation for future research. The methods section presents a guideline of how the presented research was conducted, with a step-by-step guide on how this study was conducted. This section concludes with a table that summarises the facial expression information that was used for reference to interpret the results of this study.

Section Five reports on the findings of this study and presents an in-depth analysis of how the results were approached. This section presents all of the acquired data from this study's research structured in a way that best answers the guiding research questions.

Section Six concludes this study by restating the research aim and focusing on the research questions and objectives of the study. The section also gives a brief summary of the study, followed by a discussion of the findings and the implications and recommendations produced from the presented research.

1.5 Conclusion

This section provided the background information for the presented research study focused on an interaction between computer vision and virtual avatar. The section outlined the research question, aims and objectives of this study, and the rationale for the value of this research. Finally, this section laid the foundation of information for the concepts that this study will be focusing on to answer the presented research question.

SECTION TWO: THEORETICAL FOUNDATION

2.1 Introduction

This section introduces the theories and ideas which the presented research was built upon to create a foundation of knowledge. The field of computer vision is vast and there are several concepts in this field of knowledge which need to be addressed to create an understanding of the ideas that will be used throughout this research study. This section will be introducing these concepts through a discussion of research surrounding computer vision, facial recognition, facial expression recognition, artificial intelligence, deep learning, and the methods and approaches surrounding these concepts.

2.2 Computer Vision and Facial Expressions

The study of facial expression recognition can be traced back to the 1960s when Bledsoe, Chan, and Bisson developed the first algorithm for facial recognition (Magdin and Prikler, 2017: 2). Their research was later built upon by Goldstein, Harmon, and Lesk in their development of the facial feature extraction area of facial recognition (Pasquale, Percannella and Vento, 2014). The algorithm they used was able to automatically evaluate 16 different parameters on the face by comparing data received through using a real human's input. While a breakthrough at the time, they were only able to produce a success range of 45%–75% when testing if the system could correctly identify facial feature parameters (Pasquale et al., 2014).

Research studies later spawned several detection methods that could be used in the classification stage of the recognition process. In 2002, a classification consisting of several methods was developed by Yang (Solanki and Pittalia, 2016: 21). These classification methods included: feature invariant approaches, knowledge-based methods, appearance-based methods, and template matching methods. Algorithms using the feature invariant approaches were used by researchers such as Vezhnevets, Lakshmi and Kulkarni, as they studied how skin colour could be used to improve detection accuracy while also using the grayscale edge of the input image to specify facial positioning (Magdin, Benko, and Koprda, 2019: 3). Not long after, researchers like Ghimire, Lee and Chavhan, proposed a new algorithm that used a histogram, or "snap technique", to improve image input, greatly increasing the effectiveness of image pre-processing (Magdin et al., 2019: 3).

This method was used in combination with the skin colour and edge information to improve how quickly the system is able to detect facial features while also being able to differentiate between individual faces and features. Template matching has been a widely used method through facial recognition studies, with the oldest algorithm being proposed in 1969 (Sakai, Nagao, and Fujibayashi, 1969: 234). This algorithm used several sub-templates to extract the eyes, nose, mouth, and face to create a replica of the face through a virtual model. As technology advanced to better support larger data structures, these templates were altered to use predefined patterns that were defined by researchers in advance. Appearance-based methods are grounded in the use of template matching methods, but to identify and recognize individual areas on the face, the use of machine learning (otherwise known as a support vector machine) methods provided optimal results (Magdin et al., 2019: 3).

Over the last century, several different models have been developed for emotion classification (Dyck, Winbeck, Leiberg, Chen, Gur and Mathiak, 2008: 2). These models range from a focus on universally experienced basic emotions to unique and complex emotions that need to be defined from a psychological perspective to be understood. There are, however, two models that are most prominent in the study of emotion recognition, with these models being: Ekman's "basic classification of six emotional states" and Russel's "circumplex model of emotions" (Magdin and Prikler, 2017: 2). While Ekman's classification system separates emotions into six basic emotions, Russel's model indicates that the emotional state of an individual is a dynamic system of patterns (Magdin and Prikler, 2017: 2). As an example, the expression of fear on the face consists of the widening of the pupils while contracting the muscles around the mouth, while the expression of joy consists of the reduction in pupil movement and a substantial change in the shape of muscles around the mouth. These patterns will be expanded upon further in **Section 3**.

Erkman's model would view these expressions as individual cases, whereas Russel's circumplex model would suggest that there would always be an overlap of certain features that could uniquely classify a given type of emotion (for example, happiness and surprise would have similar aspects) (Magdin and Prikler, 2017: 3). While there are several models for emotion classification, studies in human expressions have indicated a general agreement that to classify various emotional states accurately it would be necessary to recognize the physiological processes that are involved in emotion communication. As such, developments for emotion recognition have focused on behavior and physiological aspects (Machado, Beutler and Greenberg, 1999: 42). Current face analysis systems, which are able to determine the emotional state of an individual from the facial expressions analysis, operate in three basic phases, as defined by Kanade (Magdin and Prikler, 2017: 2):

1. Face detection phase
2. Feature extraction
3. Classification of emotions according to the selected model.

These steps will be referred to and used as a framework throughout this thesis, so it is necessary to understand the history of these steps and how they reached the advanced stage presented today.

Step 1: Detection: The detection phase is when the machine undertakes the task of finding an object of focus, which in this case is the features and areas of a face (Lozano-Monasor et al., 2014: 151). The amount of data that is necessary for the detection methods to be effective is quite large, as they need a surplus of reference data to learn from, so a common approach to compiling the data is to reduce the dimensions of the detected area to make the collected data considerably smaller in size (Magdin et al., 2019: 2). This enables higher computational efficiency, as less machine power is needed, while also achieving a higher success rate in facial feature detection. The most prominent techniques used in the reduction of dimensions are: AdaBoost (Viola-Jones detector) algorithm, S-AdaBoost algorithm (Jian and Loe, 2003), FloatBoost algorithm (Li and Zhang, 2004), hidden mark model, Bayes classification, support vector machines (SVM) (Magdin et al., 2019), and a general use of neural networks. Most of these methods are used in areas other than face recognition, such as pose recognition, and these techniques are able to be used either in the classification phase or in the extraction phase.

Step 2: Extraction: The extraction phase has the machine isolate and analyse individual areas of the face that was detected in step 1 and uses the extracted information for analysis (Lozano-Monasor et al., 2014: 151). The beginning of the 1990s brought about the first examples of face detection methods which used several new algorithms and techniques for information extraction (Magdin et al., 2019: 3). This method contained five basic techniques: 1. principal component analysis, 2. neural network-based methods, 3. active shape modelling, 4. active appearance model, 5. Gabor wavelet (Magdin et al., 2019: 2). Gabor wavelets, or Gabor filters, became a highly questioned topic of research as Gabor filters are used to describe the emotion being detected through the changes of the individual facial features, such as the change in the position of the eyebrows or mouth (Tian, Kanade, and Cohn, 2000: 144). The use of Gabor filters resulted in the later creation of a 'virtual mask' composed of action points (Magdin et al., 2019: 3). These action points, also called 'feature points', form together to create a system known as the geometric model and the Active Shape Model (ASM). ASM was first used in 1988, and later, this model was improved by a research team in 1992 (Yuille, Hallinan, and Cohen, 1992). Active Shape Models are extremely effective in feature extraction as the detected feature points are continuously compared with a real individual, forcing the machine to aim to describe the detected areas as closely to the real human counterpart as possible.

Step 3: Classification: The classification of emotions is the final phase of automatic facial feature analysis where the machine will analyse the extracted features from step 1 to estimate a desired outcome, such as the emotions expressed or what the age of the face is (Lozano-Monasor et al., 2014: 152). However, the algorithms used in the classification of emotions are also commonly found to also be applied in the previous two phases (Magdin et al., 2019: 3). A typical example of this is the Hidden Markov Model (HMM), which represents the use of statistical models. Statistical models are the formal basis for the creation of facial models through probabilities as they are able to fill in the blanks of facial detection errors (Rajakumari and Senthamarai Selvi, 2015: 91). Hidden Markov Models are most commonly represented as a simple dynamic network, and they are found to be used in several aspects of the facial feature analysis process. Rajakumari applied HMM in a classification of the emotional states by measuring the distance between the eyebrow and the iris and classified the emotional state of anger, fear, happiness, and disgust by measuring this distance (Rajakumari et al., 2015: 92). Similarly, technological advancements have made it so that neural networks can be used in all three phases of the recognition process (Magdin et al., 2019: 4). An example of neural networks being used in all three steps was in a study that proposed a deep learning model focused on extracting features from the backgrounds of scenes to extract facial features (Wang, Song, Liu, Song, Wang and Pan, 2018). In this study, the authors used neural networks to train the model based on a labelled face dataset. This approach used neural networks to detect the face by training the machine to recognize faces of all shapes and sizes, and the extraction of these features were simplified by referring to the trained model idea of what features would be focal points (Wang et al. 2018). The classification phase identifies the emotional expression of a face image by comparing it with the thousands of expressions in its learnt scene dictionary.

2.3 Challenges in Computer Vision

While several methods and techniques have been developed over the last 40 years, the current software available for fully automated face analysis systems still run into several issues which cause inaccuracies (Magdin et al., 2019: 4). If the position of the face is at an extreme angle in relation to the scanning device, then certain parts of the face may be out of focus and undetectable. Having the face too close or too far from the sensor would cause a loss of facial features information in a similar sense to bad lighting in the image which darkens areas of the face. Objects that cover a certain part of the face, such as glasses or facial hair, can cause a partial loss of the information but often these features are permanent accessories to a person's face and need to be included. An issue has also been found is the differences that ethnicity, age, and sex would cause in an individual's appearance. Several researchers have adopted the belief that ethnicity (race/culture) should be included among the basic and key attributes of facial analysis due to the large amount of differences in the features of people from around the world (Wu, Yuan, Wang, Gao, and Cai, 2018: 4). Traditional machine learning methods deal with race classification with two separate steps: extracting artificially designed features to train the machine to recognise varying shapes and sizes of features, and the training of a specific classifier with these features to march with an ethnic identify (Wu et al., 2018).

Face recognition in images is one of the most challenging research issues in tracking systems due to the large number of details that would affect the detection and classification of facial features (Magdin et al., 2019: 4). One such problem is that various non-standard poses or expressions are produced when extracting individual aspects of the face. A detail as small as lighting change can be enough to lead to an incorrect expression classification. Thus, the success of tracking system methods relies heavily on the strength of the extracted areas and the ability of the software to handle both high and low-quality images (Di Fiore et al., 2014). In terms of computational power: the reduction in data size and the functions in the face recognition process are essential and researchers have recently been focused on the use of modern neural networks for automated analysis (Dupré, Krumhuber, Küster, and McKeown, 2019). These challenges are dealt with by using the previously mentioned facial recognition approaches to increase the overall success of facial recognition and emotional expression classification.

These aforementioned challenges that are faced by computer vision programs can be mitigated through the use of virtual avatars to further decrease the potential problems faced by computer vision. By using a virtual environment, researchers are able to freely manipulate all factors involved in the facial recognition process, making an ideal testing environment. This study aims to use virtual avatars to test the expression recognition processes of computer vision by creating this ideal environment. Although, even with a perfect environment it would be beneficial to use advanced approaches to expression recognition to produce valuable information. 'Deep learning' is the preferred approach used in this research study and will be presented in the next subsection.

2.4 Deep Learning

The development and use of deep learning techniques (deep neural networks) removed several problems that facial recognition systems faced as it became one of the most advanced methods of facial recognition processes (Magdin et al., 2019: 4). Illumination, head pose tilt and identity bias were all greatly reduced when the use of deep convolutional neural networks (DCNNs) granted face recognition the ability to learn a large library of facial features from raw images. The potential usage of deep learning is wide and rapidly developing in areas of deep learning, and predictive analysis has started to play a key role in other facets of life, such as healthcare development (Magdin et al., 2019: 5). Research to understand participants' feelings and needs if they are unable to move or communicate with the doctor in a standard way is one of the areas that DCNNs have been helpful (Rincon, Costa, Carrascosa, Novais, and Julian, 2019). Some researchers, such as Wang (et al., 2018) apply their own approach to expression recognition that is based not only on face detection but also on additional sensors which focus on other areas of human emotion expression. The additional sensor these authors added was an algorithm based on eye movement information which collected and analysed datasets of eye movement sensing.

These methods were used to specifically extracted the time-frequency motion functions of the eyes by integrating what they called "time-domain" movements of the eye, better known as jerk motions, fixation duration and pupil diameter (Magdin et al., 2019: 5). They investigated the use of two strategies: function-level fusion (eye movements that had an unconscious purpose) and decision-level fusion (eye-movements that were intentional) along with recognition experiments that aimed to detect three emotional states through movements of the eye: positive, neutral, and negative. Their studies presented an average accuracy of only 88.64% (FLF method) and 88.35% (DLF method), which leaves much room for improvement (Wang, Ly, and Zheng, 2018: 13). Methods presented in this study would fall under a category of deep learning methods that are known as 'emotional computing'. Emotional computing is the development of machines to have them recognise and react to human emotional expressions (Magdin et al., 2019: 5). Another example of emotional computing is seen in a process known as "Markov chains", which is a mathematics-based deep learning model used to classify only two emotional states: negative and positive (Chaichotchuang, Hengsadeekul, and Sriboon, 2015). The authors aimed to calculate an emotional probability that simulates the dynamic process of change in-between different emotional states. Methods such as this offer an interesting approach to emotional state classification studies as it opens up the pathway to additional creative methods for emotional classification.

The introduction of the "mathematical model" is one of the most important developments of human-computer interaction and it is viewed to be a leading factor in what led to the current software advancements of face analysis and the classification of emotional states (Magdin et al., 2019: 6). An important aspect of the mathematical model is that it aims to help humans understand how to determine correctly, and better express natural human emotions through machines. There are several available mathematical models, such as the above-mentioned Markov chains, but the deep learning models are unique as they have the ability to overcome the disadvantages of other traditional deep learning algorithms (Bouchra, Aouatif, Mohammed, and Nabil, 2018: 23). Deep learning is able to be so effective as it is the most powerful option of machine learning, so powerful that they are found in several facets of modern society: speech recognition, computer vision, pattern recognition, and several others (Magdin et al., 2019: 5).

Deep learning models are designed to be "convolutional neural networks" (CNNs), which can be simply understood as what is used by machines to understand pictures through patterns (Kuo, Lai, and Sarkis, 2018: 2234). Convolutional neural networks are mainly used in the extraction and classification phases of the emotion recognition process due to these steps focusing on finding the important areas of the face. The CNN architectures for facial expression recognition used today were originally proposed by a research team in 2018 (Kuo et al., 2018). Their approach focused on the frame-to-sequence aspect of deep learning and successfully improved the accuracies of emotion recognition through tests done on benchmarking databases. For the extraction phase, restricted Boltzmann machines found that the use of deep learning models that were able to reduce computational complexity improved both the pace of extraction from the point of computer vision and reduced the strain on the hardware of the machine used for testing (Magdin et al., 2019: 6). Due to their noted success, Boltzmann machines and CNNs became the preferred techniques and frameworks used for deep learning when focusing on classification methods (Magdin et al., 2019: 6).

2.5 Challenges in Deep Learning

While DLMs are powerful and were proven to be accurate in facial recognition, they are not without fault. Part of the current issues with deep learning models was found in the classification step of expression recognition as they showed an inability to capture and correctly classify the change of emotional states in a face that dynamically changes its expressions (Dupré et al., 2019: 6). An example of a dynamically changing face would be a live video of a subject that is changing emotional states constantly, where a traditional example would be a static image of only one expression being displayed. Research has also shown that a shortcoming of deep neural networks (DNN) is that the classification phase's accuracy is significantly lower when the neural network used is trained with one database and tested with images from a different database (Bazrafkan, Nedelcu, Filipczuk, and Corcoran, 2017: 17). With these issues in mind, researchers aimed to develop a DNN model that has better recognition and classification capabilities for dynamic emotional states detected from uploaded video files. The FURIA algorithm was successful in this task and made it possible to use real-time data from a web camera and classify the subtle emotional facial expressions provided in dynamic expression (Li and Deng, 2018).

2.6 Conclusion

This section set out to create a foundation of knowledge surrounding the theories and concepts of computer vision. This was done through a focus on the histories and developments of computer vision, which led to discussions of deep learning methods and the advancement of the ability of machines to interpret human emotions. What was also found through this investigation was that there are challenges that are still faced by computer vision which largely stem from the limitations of input for the program, such as the quality of lighting environment greatly affecting the ability of a machine to recognise parts of the face. In this exploration of limitations, the proposed solution was that a virtual environment could mitigate the problems faced by computer vision as virtual environments can be artificially constructed with perfect conditions for computer vision programs. While both computer vision and virtual avatars are developing exponentially and becoming more a part of daily life, what this investigation also discovered was that there are few examples of interactions between virtual avatars and computer vision, which will be further addressed in **Section Three**.

SECTION THREE: FACIAL ANIMATION & EXPRESSION RECOGNITION

3.1 Introduction

This section aims to discuss the focal points of the presented research by showcasing the current field of knowledge surrounding virtual avatars and facial expression recognition. The aim of this section is to further emphasise the value that this research study presents by investigating an interaction between virtual avatars and computer vision. The term ‘avatar’ is defined in contemporary terminology as an alter ego of a user that can interact with other virtual avatars within a virtual space (Di Fiore et al. 2008: 5). Technological advancements have made it so virtual avatars have become advanced enough to be frequently used in the fields of video games, films, advertisements, medical practices, and crime scene investigation (Bernal and Maes, 2017: 3). As virtual avatars become a bigger part of our daily lives, research and development of virtual avatars have aimed to investigate how they can best imitate users in real time. One such example of the advancements achieved through this research focus is the reaching of an advanced level of realism for the virtual avatars only possible by depicting human movements by imitating bone and muscle structures in their animations (Di Fiore et al., 2008: 2).

3.2 Facial Animation Methods

Methods within computer graphics, mainly those which have been used to model the human face, have been in effect since the 1980s (Parke, 1974: 451). Approaches to researching the modelling methods in this field range from cartoon-graphic real-time animation models (Pasquariello and Pelachaud, 2001) to photo-realistic animations requiring intensive pre-rendering and powerful machines (MacDorman, Green, Ho, & Koch, 2009). Generally, two method groups can be identified: parameter-based models and muscle-based models (Di Fiore et al. 2008: 4). Parameter-based models focus less on looking at the physiological basis of the human face to create expressions and aimed to create artificial expressions through inputting specific parameters, while muscle-based models aim to emulate human muscle and skin behaviour in a realistic fashion. Both of these methods are important to highlight to understand how the programs and methods used for this research study were designed and built.

3.2.1 Parameter-based models

The first identifiable use of parameter-based systems was found in a study by Parker (1982) and then later by Magnenat-Thalmann, Primeau, and Thalmann (1987). Both of these studies used a set of parameters, otherwise referred to as guidelines or limitations, to influence the overall appearance of the animation model by editing specific areas of the face, such as the length and width, how wide the mouth opens, and how high the eyebrows are able to raise. These methods used a combination of geometry and texture morphing techniques to give the impression of facial movement and expressions (Fabri, 2006: 29). The virtual avatars in these studies were not animated in real time, rather they were modelled to what the researchers felt was a portrayal of the intended expressions which would classify these avatars as actually having posed emotional expressions. A related technique is that of “pseudo muscle animation” which ignores the anatomy of the face and only a graphical layer of skin is animated and deformed (Noh and Neumann, 1999: 101). As in the parameter-based approach, muscle groups and their interactions are not considered and would often be seen as abnormal if there are facial feature movements that do not corelate with what is understood as natural human muscle movement. As *FaceReader* is built to interpret the intricacies of a real human face, it would be best to use another method that more accurately imitates human emotional expressions.

3.2.2 Muscle-based models

Muscle-based systems were first found in a study done by Platt and Badler (1981) where they used geometric deformation operators to control a large number of artificial muscle units on a virtual avatar. Their methods were later refined and developed further by adding modelling methods that were able to have the avatar imitate more natural feeling anatomical facial muscles and there was an additional elastic nature of layers to the virtual human skin, which resulted in avatars with a dynamic muscle model (Fabri, 2006: 30). Studies as early as 1987 were also able to animate human emotions (such as anger and sadness) what tests showed to be an accurate way (Waters, 1987). This accuracy was achieved by developing a 47- muscle version of a facial model, which was also advanced enough to be used in the 1988 Pixar movie “Tin Toy” and would be considered a huge step in the quality of the available materials (Fabri, 2006: 30). While a modelling of a face that is focused on a realistic anatomical structure of a real human face will produces highly realistic expressions and animations, muscle-based models can be more accurate than parameter-based model (Bui, Heylen, & Nijholt, 2003). Muscle based models are also notably more complex and put a larger strain on the machine’s hardware, making them less useful for real-time expression recognition until the start of the 2010s when computer technology became more accessible (Park, Kim, and Whang, 2021: 3).

3.2.3 The MPEG-4 standard

MPEG-4 (2004) is an international standard created to standardise multimedia communication and it describes the modelling and animation of every aspect of virtual humans (Fabri, 2006: 30). MPEG-4 is more advanced than the previous options, as when modelling the face, it will define 68 Facial Animation Parameters (FAP) designed to guide the animation of the facial features (Garchery & Magnenat-Thalmann, 2001: 40). These parameters allow the virtual avatar to reproduce high-level expressions as they provide reference points on a facial mesh connected to a real human counterpart. This mesh is created from the measured distances between major facial features on the facial input in an intentional neutral expression state (Fabri, 2006: 31). These methods are very similar to the parameter-based animation approach described above. Although, it was found that MPEG-4 methods do not specify any particular way of achieving facial mesh deformation, it only provides a set of parameters on the face, and that MPEG-4 feature points by themselves actually do not provide sufficient quality of face reproduction (Garchery & Magnenat-Thalmann, 2001). For these reasons, MPEG-4 would not be the best choice for this study, rather the muscle-based models would likely provide the most accurate and reliable facial expression results.

The above methods to approach facial recognition are important to note as they provide a background of information on the development of computer vision and how we reached the level of complexity presented in this study. While the MPEG-4 standard is what the program used in this research project was built on, as laid out in Section 4, the influence that the parameter-based model and muscle-based models were pivotal to reach this level of complexity. By identifying the previous iterations of facial recognition processes, this research study also aims to further understand the field of studies around virtual avatars that will be laid out in the following subsections.

3.3 Computer Vision and Virtual Avatars

This subsection outlines the current methods and approaches to research in the use of computer vision in expression recognition and how they would be used with virtual avatars. Throughout the 21st century, there has also been significant research done which aimed to make virtual avatars more expressive and more accurate at portraying emotions in real-time to a similar degree to that of a human face (Park et al., 2021; Song, Shi, and Reed, 2020). Studies done in the year 2000 applied a simplified version of Waters' above-mentioned muscle-based facial animation model, by having professional actors control the eyebrow and mouth shape of their virtual avatar counterparts (Slater, Howell, Steed, Pertaub, and Garau, 2000). As the participants were professional actors, they had a deep understanding of how to portray exaggerated emotions to an audience. The study was conducted by having these participants use the virtual avatars in a theatre performance in-front of a live audience with the aim being that four emotions, happiness, anger, surprise, and sadness, were being presented through the virtual avatars (Slater et al., 2000: 104). Their findings exclaimed that the expressions portrayed were accurate representations of human emotions, although their effects on the crowd's experience could not be commented on due to the lack of a control group to compare their results to, so the true effective range of these virtual avatars was not found.

3.3.1 Facial Capture Points

Another study (Wei, Zhu, Yin, and Ji, 2004) tracked the facial expressions of a human participant and transferred these expressions onto an avatar's face in real time through an algorithm of 22 feature capture points. The expressions being detected were unknown to the system, which meant that avatar accuracy and expression recognition ability depended on three factors: How intense or distinct the user's expressions were, the quality of the input (such as the picture quality of the image of the face), and the quality of the algorithm which was used to transfer the detected facial features to be interpreted (Wei et al. 2004: 3). As this experiment was over a 15 years ago, the method did not work reliably due to constraints in computing speed and the lack of complex environments, but if redone with modern advancements the results would likely improve in reliability. Even though this experiment is old, the factors that were focused on are important even today. Intensity of the emotional expressions, image and video input quality, and the complexity of the algorithm, all have a large effect on facial expression recognition. Choosing a program that is able to reliably provide these factors would be essential for this study.

3.3.2 Embodied Conversational Agents

Another area of research that was found in the field of virtual avatars is the study of what are known as Embodied Conversational Agents (ECA) (Fabri, 2006: 21). These avatars are not controlled by a user, rather these are virtual characters that react and portray expressions through the use of a behavioural model to control their actions. An ECA requires a well-developed emotion-appearance mapping system to be expressive and believable to human participants, which helped develop the use of virtual avatars as a whole (Bailenson and Blascovitch, 2004). A study done by Bartneck (2002) developed an autonomous virtual character known as eMuu, shown in **Figure 1**. eMuu was created with the aim of achieving a high level of expression accuracy whilst using an extremely basic animation model and minimal facial features. eMuu was a very simple avatar yet research showed it to effectively express three emotions, Anger, Sadness and Happiness. This ability was granted through only the use of stylised representations of eyebrows and a mouth, with only a single cyclops-like eye. While eMuu was not designed to represent real people, as its range of expressions would be insufficient for realistic expressions, this research highlighted the ability of virtual avatars to portray expressions through only minimal facial features (Bartneck, 2002). Having confirmation that a virtual avatar as simple as eMuu being successful in expressing emotions leads one to consider how this accuracy compares when using a machine, rather than a human, to interpret the emotional expressions.

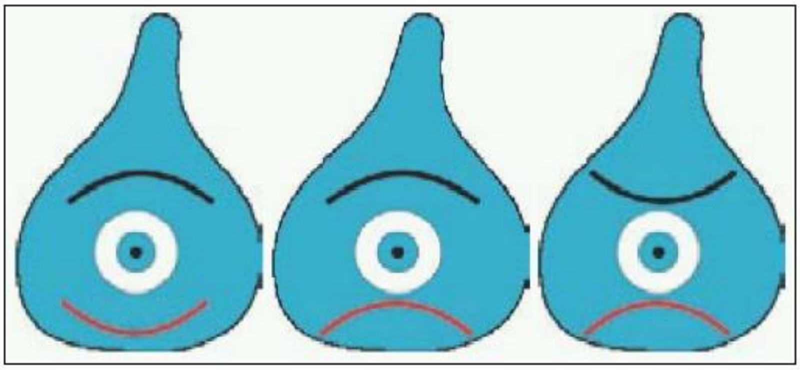


Figure 1 eMuu avatar which shows how simple an effective avatar can be Bartneck (2002)

3.3.3 Virtual Avatars & Non-verbal Communication

Other noteworthy studies surround the effect of virtual avatars on non-verbal communication and the use of virtual avatar behaviours that are able to improve the emotional expressions of virtual avatars (Fabri, 2006: 31). Manninen and Kujanpää (2002) studied how virtual avatars would have an effect on several aspects of non-verbal communication. While they did not focus on the diversity and dynamism of facial expressions, they found that giving their participants the ability to change the expressions on a virtual avatar resulted in the participants' having an increased intent to participate in non-verbal communication (Manninen and Kujanpää, 2002: 392). They were also found to display and respond with facial expressions in a more purposeful, yet playful way, meaning the emotions and expressions displayed were more posed and less natural yet noticeably more frequent. This concept was taken further when Garau (2003) investigated the relationship between visual and behavioural quality by studying the use of eye-gaze on making an otherwise non-expressive virtual avatar feel more expressive to participants. While Garau (2003) used human-controlled avatars, other researchers (Vinayagamoorthy, Garau, Steed, and Slater, 2004) looked at using an Embodied Conversational Agent. Garau (2003) found that an avatar that has eyes which move in a relative motion to the conversation provided a marked improvement in terms of conversational quality when compared with an avatar that only exhibited feature movements. Additionally, Vinayamoorhy (et al., 2004) further concluded that there is a requirement for consistency between a character's visual realism and the behavioural realism to achieve a natural feeling to virtual avatars. What these studies focus on is the human aspect of interaction with these virtual avatars but fail to address whether or not these factors change the way computer vision will identify the emotional expressions.

The common theme in the above studies is that there has always been human participation in some form to understand the emotional expressions. Researchers would artificially create movements into posed-emotional expressions (ingenuine), or a human participant would be used to analyse the virtual avatar and determine which emotional expressions are being displayed. The presented research aims to mitigate the human involvement in the expression analysis and virtual avatar animations to properly investigate the interactions between computer vision analysis and virtual avatars. The aspect of analysis that will be heavily human involved will be the understanding of why certain facial feature shapes would be able to produce different emotional expressions as this would potentially create a foundation of conversation for future studies. This analysis will be done by the researcher (Mr. van Rooyen) by referring to the current understanding of how emotional expressions are presented in the face as presented above. The hope in this analysis is that this research would also be useful in other aspects of daily life, such as improving the design of virtual avatars for intentional emotional expression-based interactions.

3.4 Understanding Facial Emotional Expressions

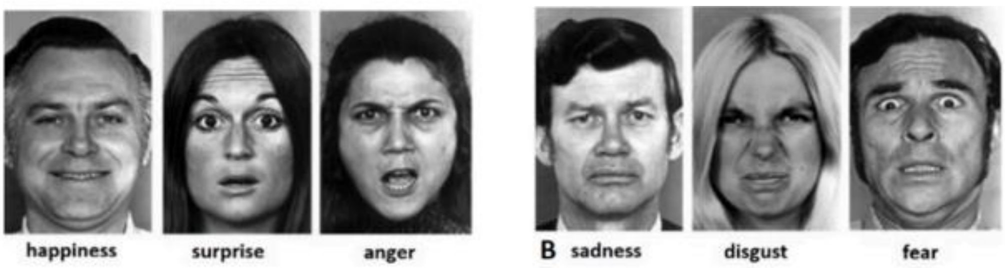


Figure 2 Examples of the six basic emotions portrayed on a human face (Geraets et al. 2021)

The following sub-section is a summary of information from research papers done by Ekman and Friesen (1975, 1982, 1986), Fabri (2006), and Dupré (et al. 2019) where they outlined the six basic emotional expressions, as seen in **Figure 2**, and how they are portrayed through the face. All information in this section has been adapted from these authors to fully understand how and why these emotions are expressed. One needs to understand the situations that would cause the emotion to be felt, as the expressions corelate with behavioural reactions to these situations, to fully analyse the facial features used in the presented research. Situations that cause emotional expressions would also effect how intense the emotion is, how long the expression is visible for, and the expression would be similar to a reaction to a situation as a survival response.

As an example, "surprise" affects how wide the eye opens, allowing access to more visual information to the brain, while disgust would contract the eyes and nasal area as if the body is attempting to move away from or hide from what is causing the disgust. An outline of the known causes of emotional expressions, and the subsequent physical manifestations of these expressions, are detailed below and will be used later in this study as a reference to understand why certain facial features produce different emotional expressions.

3.4.1 Happy

Happiness is known as the only positive emotion in basic emotion theory, and Ekman (1970) identified several paths which would lead to happiness in human expressions:

> *Pleasure* (physical): Humans tend to feel happy in when pleasurable sensations are expected, or they would be happy after experiencing a pleasurable situation.
>
> *Excitement*: Happiness can come with an expectation of something exciting or interesting, especially if the intent is to relieve a state of boredom.
>
> *Relief*: People feel happy when a negative emotion, such as anger or sadness, ends.
>
> *Affirmation*: Happiness is felt when a person's work is acknowledged or when their view of themselves is enhanced by an outside source.

Happiness often varies in both type and intensity. People can be mildly happy, or experience ecstasy. Happiness can be expressed silently or audibly, and the facial expressions shown during happiness will vary from a light grin to an open mouth to produce laughter. Happiness is primarily shown in the lower region of the face as the intensity of happiness is differentiated by the positioning of the lips and how wide open the mouth is. While there is no distinctive brow or forehead movement shown when experiencing happiness, the lower eyelid may show wrinkles going outward from the outer corners of the eyes (known as ‘crow's-feet’), but they are not classified as being tensed. The mouth shows happiness through the corners of the lips being drawn back and up. Intense happiness may cause a parting of the mouth (i.e., Laughter) which is measured by looking for the exposing of the teeth and the raising of the cheeks.

3.4.2 Sad

Most often people are sad about losses such as: death, rejection, loss of health, or loss of an opportunity. Disappointment and hopelessness can also cause sad feelings as expectations are not met, but the realisation is not as sudden as fear or surprise. Sadness varies in intensity from slight feelings of gloom to the extreme sadness felt during mourning. While the appearance of most of the other basic emotions becomes more distinctive when the intensity of the emotion increases, this is not the case for sadness. In its most extreme form, there may be no facial clue to sadness other than a general loss of muscle tone in the face. Primarily to display sadness one would look at the lips, cheeks, and lower jaw, however there are other distinctive facial clues in cases of less severe emotional intensity.

- The inner corners of the eyebrows are raised and may be drawn together as the skin below the eyebrow is triangulated with the inner corners moving upwards.
- The inner corner of the upper eyelid is raised, with the outer corner dropping lower.
- The raising of the lower eyelid is not as common but can increase the sadness expression, this was also found to be intentional.
- The corners of the lips are dropped down, and the lip may be trembling in more intense feelings of sadness.

The focus is mainly on the upper area of the face as what is known as ‘sadness eyebrows’ and ‘sadness forehead’ are not as easily covered as they are reflexive muscle reactions, not intentional. The mouth or nasal area can also tend to show anger or happiness along with the sad upper facial area, producing a blend of sadness-anger, and sadness-happiness respectively. Both blends are found to be a result of a person attempting to mask their expression of sadness by expressing anger or happiness through movements of the mouth.

3.4.3 Anger

Anger is an aggrieve emotion that is expressed to signal aggression to those around the person feeling angry. Anger is often accompanied by a rise in blood pressure and the reddening of the face which are both caused by the fight-or-flight response and an increased heart rate. There may be movement observed as the person moves forward towards the cause of the anger, and anger can be caused by a number of different sources. Anger has been known to be caused from frustration with another subject, intentional provocation, psychological pain, and people have been found to even become angry through just the presence of another person expressing anger. Anger is a highly variable expression with several intensity levels and has been observed to build gradually from a low intensity (irritation) into a more intense anger (rage), but it is also possible for intense anger to be felt.

The facial feature movements which are commonly associated with anger revolve around the mid-section of the face:

- The eyebrows are lowered and drawn together sharply.
- There are two variants of anger expressed through the eyelids as the upper and lower eyelid are shown to be tense, with the upper eyelid also being lowered by the movements of the eyebrows.
- The lower eyelid may or may not be raised and the iris stares out in a direct fashion towards the cause of the anger.

The lips are more varied as they would change depending on if the person is expressing their anger physically or verbally. This is seen in one of either two basic positions: pressed firmly together with the corners moving down, this is commonly associated with physical violence situations, or they may be open and tensed in a 'squarish' shape, which is associated with verbal confrontation.

3.4.4 Surprised

Surprise is the briefest emotion as it usually does not remain as a facial expression unless a person is experiencing a sequence of surprising event. Surprise is triggered by unexpected events, but it is not considered as only a negative emotion. Surprise can be felt as a positive experience (such as surprise gift or compliment) or a negative experience (unexpected loss of a loved one). Surprise may also arise when reality does not corelate to what was expected of an event (such as plot-twists) and this expression will only last until the surprised individual has evaluated what event has occurred. Once the reason for surprise is determined, the emotion would fade away and be replaced by another emotional expression.

When surprise is being expressed:

- The eyebrows are raised higher on the face and curve with their centre being raised higher than the corners of the eyebrow.
- The eyes open wide, with the lower eyelids looking relaxed and the upper eyelids raising higher with a more intense expression.
- The opening of the eyelids also exposes the white of the eye above the iris.
- The individual's jaw would drop open so that the lips and teeth are parted, with the opened mouth seeming to be relaxed rather than tense.

Determining the intensity of surprise can be done through the lower face and examining the extent of how far the jaw has dropped. Once a person does not feel surprised, the emotion will be replaced with another expression with remnants of surprise fading quickly.

3.4.5 Scared

When experiencing the anticipation of a situation that would cause psychological or physical harm, which may be real or even imagined, a person feels fear. The anticipation of harm is often more impactful on human expression than the actual pain itself and fear can also be experienced without any anticipation of harm when a person is caught completely unaware of a developing situation. This is similar to 'Surprise' but there are noteworthy differences between fear and surprise as while surprise can be pleasant or unpleasant, fear is always an unpleasant expression. Additionally, one can feel fear of an event they know of, whereas one cannot be surprised at an event that is known. The length of fear and surprise are also a differentiator as surprise fades after the event is experienced and interpreted, while fear can continue after the event was experienced and understood.

Some of the facial feature movements commonly associated with fear are best described in comparison with surprise:

- The eyebrows are raised and be drawn together, with the inner corners of the eyebrows being closer together than when expressing surprise.
- The eyes are opened and tense, with the upper eyelid being raised and the lower eyelid tensing and moving whereas the lower eyelid is relaxed in surprise.
- The mouth is open, and the lips are either tensed or stretched outwards and it may also be drawn back tightly, as opposed to the open but relaxed surprise mouth.

When fear varies in its intensity, these differences are reflected mainly in the mouth and eyelids of the person expressing fear. There is an increased stretching and opening of the mouth when the feeling of fear increases, and an even greater intensity of fear would cause a rise in the upper eyelid and a severe tensing of the lower eyelid as the eye seems to contract.

3.4.6 Disgusted

Disgust is described as a feeling of avoidance and aversion to external sources that are not pleasurable to our senses. The thought or sight alone of something that would be unpleasant to the senses of taste, touch or smell can cause the expression of disgust in a reflexive fashion. The tastes, smells, and texture that are disgusting will vary from person to person and it is important to note that what is repulsive to one person may be attractive to another, causing a wide variation in results when disgust is the focal point of a study. Expressing disgust usually involves the face and body moving away from the cause of disgust with the most important clues to this emotion being in the mouth and nasal area.

Disgust is most associated with:

- A contraction of the face as the facial muscles moves inwards towards the centre of the facial area.
- The eyebrows lowering, which would also cause the lowering of the upper eyelid as well.
- The face will wrinkle below the lower eyelid as the face contracts.
- The upper lip raising, and the positioning of the lower lip will vary as it can be either raised and pushed up towards the upper lip or it will be lowered and start protruding outwards.
- The nasal area wrinkles, with the more intense feelings of disgust causing more wrinkles and a raising of the cheeks.

Contempt is an emotion very similar to disgust, and while it is not a focal point of this study it is mentioned as it is very similar to disgust, but contempt is typically experienced when the source of displeasure are people or the actions of people rather than the source being tastes, smells, or touches. In contempt, there is an element of condescension toward the source and people feeling contemptuous toward other people, or their behaviour, often feel superior in some form or fashion. Unlike with disgust, there is not necessarily a need to get away from the cause, so the contractions of the facial features during contempt are less apparent and should not be confused with disgust.

3.5 Conclusion

This section presented a discussion on the methods and approaches to modern day computer vision and virtual avatars. A disconnect between the two areas was discovered as there is little information available which focuses on how computer vision and virtual avatars would interact. The use of human participants in these fields of research have been necessary in the past, but as the technology around virtual avatars and facial recognition grow in both use and complexity, a shift to machine-only interactions may give insight into future developments and how these fields of study are intertwined. The presented research aims to provide information on this interaction by investigating the use of expression recognition on 3D virtual avatars. The results of the expression recognition experiment will be rationalised and framed through the emotional expression information presented in this section. The processes for this study are laid out fully in **Section Four**, with a discussion on the value and use of these findings for future research.

SECTION FOUR: METHODS & MATERIALS

4.1 Introduction

The approach to this study was built on the research outlined in **Section Three**. This section will begin by giving a brief overview of the procedure of this experiment, followed by the materials used and reasoning for the choices of materials, and finally there will be a step-by-step breakdown of the methods used for recreation of this study. It was necessary to do a practical experiment for the presented research as the current field of knowledge surrounding computer vision and virtual avatar interactions is extremely limited. It was also noted that the amount of available facial recognition programs and virtual avatar libraries would mean that it is necessary to scope-in to a specific set of examples to create valuable results. By approaching the ideas presented above with an experiment of my own, I will be able to refer to specific examples that can potentially create value for further research studies in these fields.

4.2 Materials

Section Three outlined examples of the available options available for virtual avatar animation, computer vision software, and expression recognition. Through studying the available options, the following materials were chosen for this research project:

Facial Expression Recognition: *FaceReader*

Virtual Avatar Creation and Animation: *Animaze*

Peripherals: *AlterCam*

These programs were chosen for various reasons, with these reasons being outlined below, but the main factor of choice is that they each provide the optimal features for this research study to allow the interaction between computer vision and virtual avatars to be tested. The specifications of the machine used in this experiment is also laid out at the end of this section to present the environment testing was conducted in.

4.2.1 FaceReader

FaceReader is a facial analysis program that was built by Noldus Information Technology (2022) to accurately detect facial expressions through an analysis of facial features. While there are several commercially available facial expression recognition programs available, *FaceReader* was chosen as it has several features that would improve the research procedures, and Noldus provides ample research that supports the accuracy of *FaceReader's* expression recognition feature (Noldus Information Technology, 2020). *FaceReader*'s features are laid out below to provide a background of support for why this program was the right choice for this research study. Additionally, permission to use FaceReader was given by Noldus, with the permissions presented in the appendix.

4.2.1.1 Choosing FaceReader

The choice to use *FaceReader* can largely be attributed to this program following Ekman's Basic Emotion theories presented in **Section Three**, *FaceReader* has been trained with the goal of classifying when a human face is expressing the emotions of happy, angry, sad, scared, disgusted, and surprised, and providing a measure of the intensity of that emotion. A neutral emotion, which is classified as a lack of emotional expression being present, is also able to be detected. FaceReader's expression recognition is conducted through three steps, as seen in **Figure 3**, which match the steps used in other computer vision programs shown in **Section Three**:

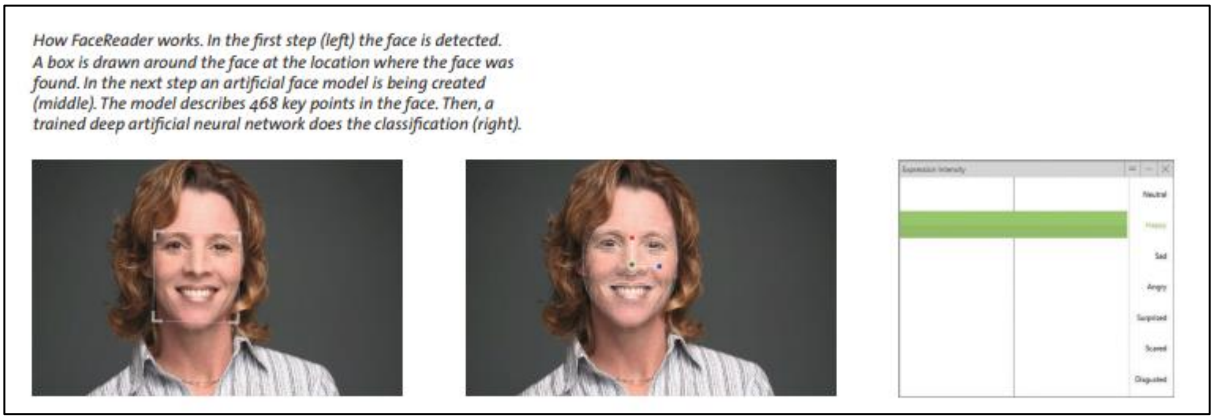


Figure 3 Three steps of facial expression recognition by FaceReader (Loijens and Krips, 2021)

Step 1: Face finding. The position of the face is found using a deep learning algorithm, trained to search for the presence of facial features by looking for the patterns that indicate a human face.

Step 2: Landmark Detection. FaceReader produces an artificial face model, which describes the location of 468 key points in the face detected in Step 1. It does a single quick estimation to find landmarks in the face, such as the eyes and mouth, before using Principal Component Analysis to compress the key points into a simple vector image of the face and its features (Loijens and Krips, 2021).

Step 3: Classification. The facial expressions are classified through the use of a deep artificial neural network, which was trained using 20.000 manually annotated images, to allow *FaceReader* to recognize patterns in the face matching facial expression patterns in its database. Once patterns are recognised, *FaceReader* outputs the emotional expression that it believes is being expressed by the detected face. The outputs of FaceReader will be discussed further below.

The use of *FaceReader* has been limited to use on real people rather than virtual avatars. This is not a shortcoming of FaceReader, rather it is important to note that the testing and training of the FaceReader model was not designed to interpret the emotional expressions of a virtual face. This is further emphasised by FaceReader's "remote photoplethysmography" module (Loijens and Krips, 2021: 11) which detects the blood pressure and reddening of the cheeks during emotional expressions. Virtual avatars do not have the ability to produce artificial reddening of the cheeks in real-time, unless they are specially designed to do so, thus this feature would not work on the average virtual avatar. This feature will NOT be used in this research study, rather the mention of this features emphasises that FaceReader was built to detect emotional expressions on human faces rather than virtual faces. This does not invalidate any results presented by *FaceReader* as the aim is not to test for expression accuracy, rather the aim is to analyse how changing virtual facial features will affect *FaceReader's* expression recognition.

4.2.1.2 Additional FaceReader Features

FaceReader hosts several other features which were not used in this research study but may be useful in future research. It is possible to add custom expressions to the software for classification, and it is also possible to detect the movement of eye gaze, or whether eyes and mouth are closed or not. This would add the possibility of behavioural realism when testing the effectiveness of virtual avatars emotional expressions. There are multiple face models available in FaceReader such as the models for East-Asian people and babies (6-24 months). Based on the TFEID [6] dataset of East-Asian faces (268 images), the East-Asian model performs with an accuracy of 93%, while the General model achieves an accuracy of 91%. This further emphasises the fact that race and ethnicity differences will have an effect on emotional expression readings, meaning the face model used in future studies would need specific training. As an example, a study that incorporate racial differences would need to use a face model that was trained using data from various races

Valence readings are also a part of *FaceReader*, which indicate whether the emotional state of the subject is positive or negative. 'Happy' is the only positive expression, while 'sad', 'angry', ‘scared’ and ‘disgusted’ are considered to be negative expressions. ‘Surprised’ can be either positive or negative so it is ignored for valence calculation. The valence is calculated using ‘positive’ minus ‘negative’, meaning the intensity of ‘happy’ minus the intensity of only the highest negative expression recorded to account for the multiple negative expressions. Valence readings would be useful in future studies which analyse the differences in ranges of expression intensity within virtual avatars, or when being compared with a human counterpart.

4.2.1.3 FaceReader Output

FaceReader is able to provide a classification of facial expressions into the six basic emotions and a neutral emotional state. These classifications are visualised in several different representations of data such as graphs, tables, charts, with the option of the results being exported to text log files as well. The data *FaceReader* provides consists of an overall average intensity of emotional expressions, which will be used throughout this study, and the classifications from individual frames from the participant video. These frames are isolated and each emotional expression that was detected in that frame is given an intensity value between 0 and 1, where ‘0’ means that there is no sign of that emotion in that frame and ‘1’ means that the emotion is not only present, but an intense representation of that emotional expression where intensity is how strongly the emotion is being expressed. The accuracy of *FaceReader’s* intensity readings were tested through comparing the analysed intensity values with intensity values annotated by humans.

The presence of facial expressions are rarely isolated as emotional expressions are often caused by a mixture of emotions. This means that it is possible that multiple expressions may be detected simultaneously with varying intensities due to some emotional expressions sharing traits (such as ‘surprise’ and ‘scared’). Being able to note the presence of several emotions in one frame makes *FaceReader* even more useful as through analysing the *FaceReader* expression results it would be possible to see which facial features tend to portray various emotional expressions simultaneously. As emotional expressions share traits, knowing which facial features are able to produce these similarities would be helpful in the design of emotionally expressive avatars. Additionally, having access to the intensity of the emotional expressions, not just knowing it is present, will assist in finding which facial features are able to show variations of them same emotional expression.

4.2.2 Animaze

Animaze is a face tracking software that allows its user to control a 3D virtual avatar through the use of a simple web camera or phone camera. This face tracking is possible through PC version of the program using *Hyperface,* a facial recognition library that is trained to specify and track the locations of several facial feature points in a similar process to the MPEG-4 standard mentioned in **Section Three**. *Animaze* animates its virtual avatars by linking the detected facial feature points to similarly placed feature points on a virtual avatar. As an example, the feature points outlining the mouth on the detected face will be linked to the feature points outlining the mouth on the virtual avatar. Animaze will then imitate the detected placements and movements of the real human face's feature points by matching the movements on the virtual avatar's feature points. This is done in real time, allowing the virtual avatar to produce dynamic movements between facial expressions.

4.2.2.1 Choosing Animaze

While the facial motion tracking is an important feature, the main factor behind choosing Animaze for this research study was its large library of facial feature options. Animaze is able to seamlessly alter and edit individual aspects of a virtual avatar, while keeping the rest of the facial features the same. One is able to choose between different options of hair, hair colour, skin tone, the clothes worn, and the accessories the avatar is wearing. For this experiment we will be editing only the six facial features options that are available: the shape of the face, the eyes, the eyebrows, the ears, the nose, and the mouth. Only these facial features will be focused on, as **Section 3** elaborated that these are the areas of the face that are focused on when interpreting facial emotional expressions. This is not to say that the colour of the skin or the areas around the face are not important, but rather this study requires a scoped-in view of only these features to answer the research question.

Animaze allows the user to import their own 3D virtual avatar to be used for facial motion capture. This process is laid out in the *Animaze* documentation and would be useful to research the variations in virtual avatars and how more, or less, realistic features would be presented and analysed by *FaceReader*. *Animaze* also provides the user with a library of dozens of virtual avatars that range from animals, to cartoon characters, to an eggplant. These variations in avatars are used in the same way the default avatar is used, but the output would produce vastly different results if presented to *FaceReader*. *FaceReader* detects and analyses facial expressions by pinpointing certain facial features, as laid out above, so when it is required to analyse a non-human face, it would not be accurate. While not useful for this study, a future study that researches the use of inhuman virtual avatars would find this feature extremely useful.

Section 3 discussed the shortcomings of facial recognition programs being lighting, the environment, and clarity of the face. *Animaze* also allows the user to edit the lighting on the face, creating shadows or removing artificial lighting, enabling further testing in future on how to influence computer vision results. The background of the avatar is also available to be edited, as the user can change the backdrop into a still image or a separate video, allowing this study to be done with a clear and neutral background to avoid external influences. As the resulting facial expressions of the virtual avatar will be documented using screen recording, and the view of the face can be zoomed in, or out, allowing the clarity of the face to not be a noteworthy issue for this research study.

4.2.2.2 Additional Animaze Features

Additionally, *Animaze* presents the user with a number of 'Advanced Tracking' settings that can be changed to alter the presentation of the virtual avatar. The smoothening of movement for the features can be increased or decreased, making them appear more uneven when decreased, the range of movement for the iris can be altered, the features can be set to move in unison in a mirror-like fashion, there are dozens of options similar to this that allow the user full control over the virtual avatar. For this research project all settings were reverted to their factory-setting state to assist in experiment recreation and to reduce the research confirmation bias that may form through my editing of the settings. It would be beneficial to research the extent of these settings effect on expression recognition in future.

4.2.2.3 Animaze Output

Animaze was originally made with the intention of being used by streamers and video content creators, so there are features that support the exporting of content from *Animaze* to a video format. The 'Screen Capture & Record' feature was used for this research project, which would create an MP4 video of the virtual avatar. The recording would start when the demo video began, and the recording would end at the same time to ensure the exported videos are all the same length. The length of the videos would be important to keep consistent as FaceReader's intensity average results can be skewed by the final expressions being shorter or longer in length. The created MP4 videos run at 30 frames per second and are of 720P quality. High quality videos are beneficial for FaceReader as they provide greater detail for analysis.

4.2.3 AlterCam

AlterCam is a program developed by **Bolide Software** where the user has access to several web-camera related features. *AlterCam* is able to add several effects to video feeds, such as filters, borders, and virtual backgrounds, but the feature of focus from this program is the 'Video Source' feature. This allows the user to create a false web-camera feed that presents a chosen video source file as a web-camera output. This feature was used to provide the demo video as the input source for *Animaze* as *Animaze* only registers a web-camera feed as an input. Additionally, it is helpful to use a pre-recorded video as the input for *Animaze* to keep the facial motion tracking consistent throughout the testing sequence. To answer the presented research questions in **Section Two** it will be necessary to compare the results of *FaceReader's* expression recognition, so accurate results will only be possible through using the exact same input throughout the testing sequence.

4.2.4 Testing Materials Summary

To allow the presented research to be recreated, the exact specifications for what was used for testing are provided below.

4.2.4.1 System:

<u>Machine used for testing:</u> Keenan van Rooyen (researcher) personal desktop computer.

<u>OS:</u> Microsoft Windows 10 Home, Version: 10.0.19042, Build 19042

<u>BaseBoard:</u> ASUSTeK COMPUTER INC., PRIME X370-PRO

<u>Processor:</u> AMD Ryzen 7 1700 Eight-Core Processor, 3200 Mhz, 8 Core(s), 16 Logical Processor(s)

<u>GPU:</u> NVIDIA GeForce GTX 1080, Driver Version: 30.0.14.9649

4.2.4.2 Programs:

<u>FaceReader:</u> Version 9.0.16, Build date: 10/8/2021, License: Provided by Noldus

<u>Animaze:</u> Version 1.25.9706, Steam Release, License: Free software, Research use allowed.

<u>AlterCam:</u> Version 6.0, Build: 3341, License: Free Trial version, Research use allowed.

4.3 Methods

The human input subject used for the presented research is found in the demo video 'facial_expressions_demo.mp4' which was provided with the *FaceReader 9* software. This video is 32.20 seconds long and runs at 20 frames-per-second. The choice to use this demo video as the input was due to the contents of the video consisting of a Caucasian male in his early 30s demonstrating the six basic emotions, as seen in **Figure 4** below. This video was used in the presented research as the video was verified to be created and tested with *FaceReader's* expression recognition feature to showcase its abilities. This also means the video was created with optimal conditions to be used by a computer vision software, such as:

- evenly lit facial features
- a portrayal of the six basic emotions
- a direct view of the subject's face
- clear view of all the facial features that are measured in expression recognition

Figure 4 below shows an example of the default *Animaze* virtual avatar as it attempts to mimic the emotional expressions of the input video. The details on Animaze and the methodology of the study to produce the imitation will be laid out later in this section.

Figure 4 Screenshots of the 'Demo Video' (FaceReader, 2022)

The demo video provided by *FaceReader* was trimmed to a 30 second time frame for this study and will be used as the input for the virtual avatar's motion capture throughout the experiment as this would keep a consistent input for the *Animaze* software. The screen recording feature was used to create a recording for all the selected facial features from *Animaze*, with only one feature being changed from the default avatar, shown in **Figure 5,** at a time. The choices for which features count as 'default' were taken from the first available option in the Animaze editor. These recordings were organised into testing groups depending on which facial feature was changed out of the shape of the face, the mouth, the eyes, the eyebrows, the nose, and the ears. The different videos were renamed as "participants", where each participant was the default avatar with a different facial feature selection. The different facial features were separated into *FaceReader* test groups depending on which facial feature was changed for that video and these test groups were then analysed using *FaceReader's* Facial Expression Recognition feature (Loijens and Krips, 2021).

The separation into test groups was done to make it easier to analyse the results and more efficiently answer the research questions. The overall aim of the presented research is to investigate how different facial features on a virtual avatar would affect FaceReader's detection of emotional expressions. The virtual avatars with varying facial features, referred to further as just 'participants', in each test group would have only slight changes to the facial structure which would make any variations in expression detections be due to the changes that those facial features alone have caused. The hypothesis is that by isolating these facial features into their test groups, it would be possible to understand how *FaceReader* analyses emotional expressions through the use of specific facial features.

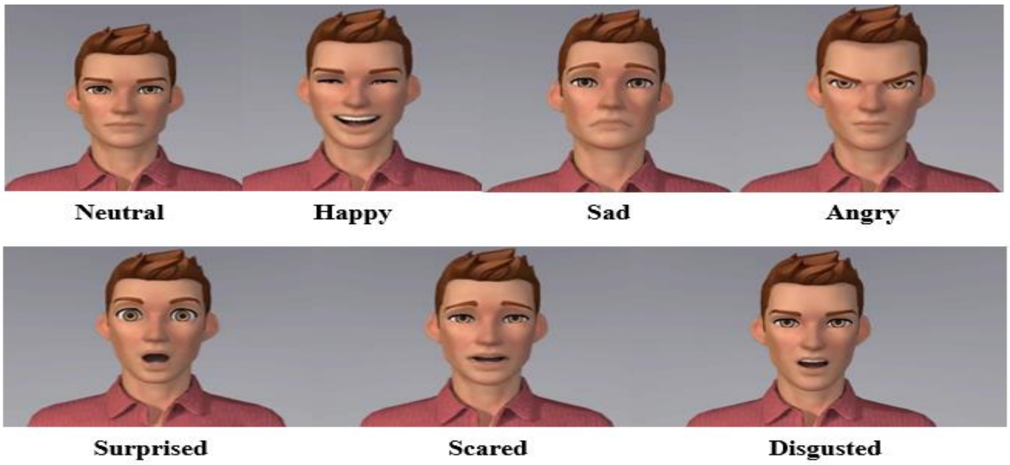


Figure 5 Screenshots of the 'default avatar' imitating the human input

Additionally, these six facial features were chosen as they were noted as the main areas of the face that are used in emotional expressions, as laid out in Section 3, with the exceptions being the 'Face Shape' and 'Ear' groups. There is no direct mention of these facial features being directly used in emotional expression, but they were used to investigate if they would have a noteworthy effect on expression recognition. In theory, they should not cause a noticeable change in expressions detected as they are not mentioned in FER documentation, but they would likely fall under environmental factors that will affect the expression recognition due to how they change the reading of other facial features.

4.3.1 *FaceReader* Calibration

Calibration is what is used to assist *FaceReader* in adjusting to each analysed subject and their facial features (Loijens and Krips, 2021). Calibration is necessary to mitigate what is called ‘personal-bias’, which may occur within certain faces as some facial features tend to express an emotion when in a neutral state due to these facial features being naturally shaped to look similar to an emotional expression. An example of this would be eyebrows that are shaped to angle sharply down and inwards, making the face look angry even when no expression is intended. Without calibration, facial expressions in *FaceReader* are known to be less balanced as each face would have personal biases towards certain emotions that, even if slight, will skew the expression recognition results. An example of how this is mitigated would be *FaceReader* recognising that the previously mentioned angry-eyebrow participant is not showing anger, rather their neutral expression leans towards anger. The resulting expression recognition would account for this by measuring slight anger as neutral.

Calibration for the presented research was conducted using the ‘continuous calibration’ option which continuously aims to calculate the average neutral expression of the subject being tested. *FaceReader* uses that average to mitigate the personal bias that may be occurring in that subject’s analysis. Calibration is vital in the presented research as there are facial features that would be heavily bias towards certain emotional expressions as their design is to be noticeably different from the other features in their category. Being aware of personal bias will be extremely important in understanding why facial features will produce varying results, and how these variations would be accounted for. Additionally, personal bias can also be beneficial in this study as it may be possible to force *FaceReader* to lean toward, or away from, identifying certain emotional expressions by using specific facial features.

4.3.2 *Animaze* Calibration

Animaze also has a calibration feature, that is optional, but will provide a more accurate imitation of the input from the web camera. When calibrating, *Animaze* will take more time to identify the facial feature points by analysing the face that it detects and should be used before the facial capture process begins. This is an optional feature as *Animaze* is built well enough to not need to calibrate for use, but the importance on this study is providing a continuously accurate representation of the inputted video. As such, this calibration option was used for every participant and every time that a new recording was made to ensure *Animaze* recognised the placement of the facial features correctly. The aim of this study was to analyse how the facial features would alter FaceReader’s expression recognition, so having a consistent output provided by Animaze is extremely important.

4.3.4 Ethical Considerations

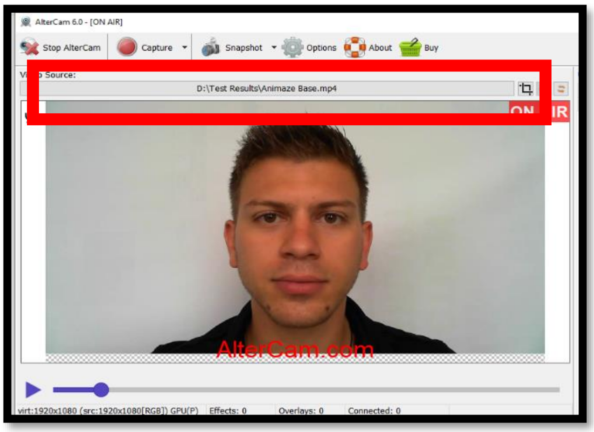
Ethical approval for this research was given by the ethics committee of the **University of the Witwatersrand Ethics Department**. No human test subjects were used, only freely available online references and images or resources provided by the mentioned parties with full permission. All access to the materials used in this research study were used with full permission given by the parent companies, with permission proof found in the appendix. *FaceReader* was provided by **Noldus Information Technology** with their permission to use the software for the presented research. There was no direct involvement with **Noldus Information Technology** when the presented research study was conducted or thereafter. *Animaze* is a free software available for download by **Holotech Studios Inc.** Interactions were made with the developers behind *Animaze* to gain insight into the program and the design philosophies behind the program. Permission was gained to use *Animaze* for this research project.

4.4 Step-by-Step Guideline

This subsection presents a guide on how the testing process was conducted to allow accurate test recreation. The associated pictures were all screenshots taken by the researcher (myself) for this study and were not considered standalone figures.

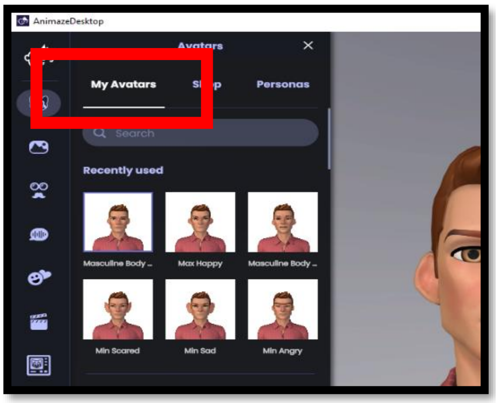
4.4.1 Set-up Guide

Step 1: The demo video is placed into *AlterCam*



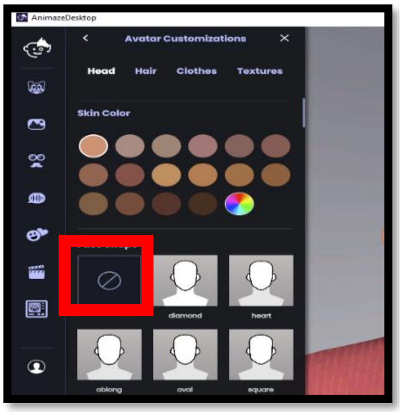
Step 2: In Animaze, the source web camera was chosen as “AlterCam Virtual Camera” to simulate the use of a web camera. This was done as Animaze does not support using pre-recorded videos natively.

<u>Step 3:</u> In Animaze, under “Avatars” > “My Avatars” the avatar named “Masculine Body Type” was chosen.

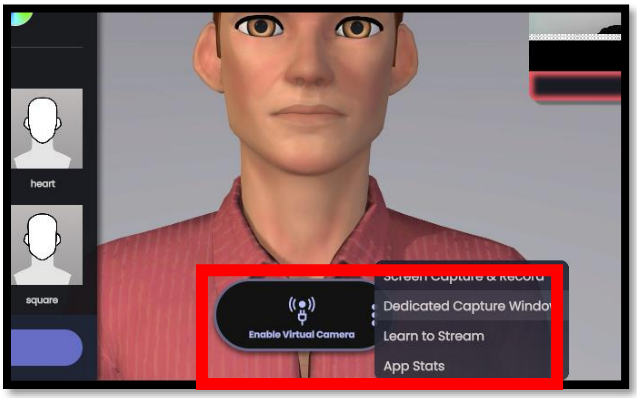


<u>Step 4:</u> The “Hair”, “Clothes” and “Textures” Menu options were left unchanged. All other settings in Animaze were left untouched from the default settings provided in a new downloaded version of the software.

<u>Step 5:</u> To achieve the “Default” avatar, under the “Head” section all skin tones and colours were set to the first option (most upper left-hand choice) and the facial features were all set to the prohibition symbol option (most upper left-hand choice). No facial hair was used.

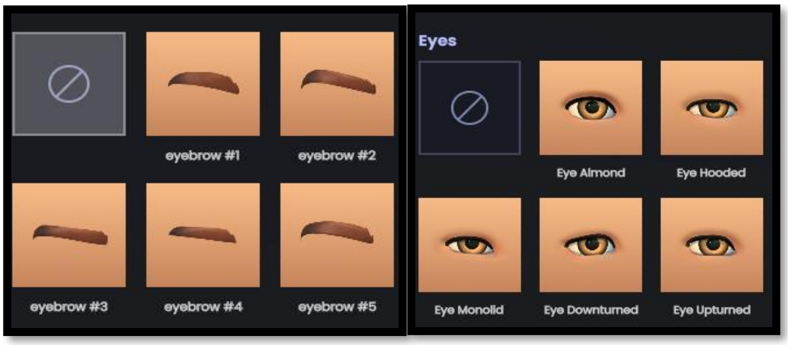


<u>Step 6:</u> To activate the screen capture setting, the three dots on the "Enable Virtual Camera" button was used.

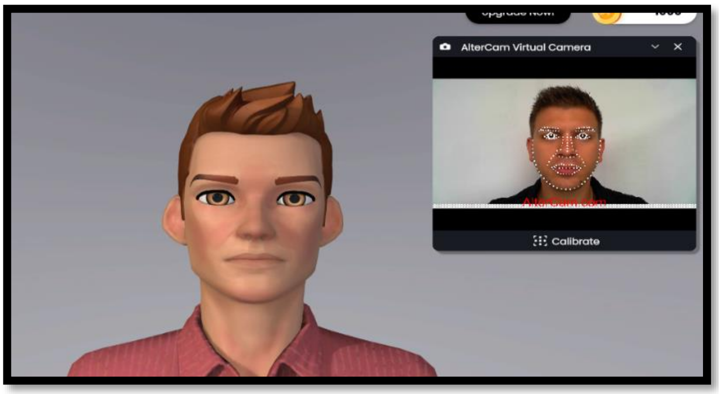


4.4.2 Testing: Animaze

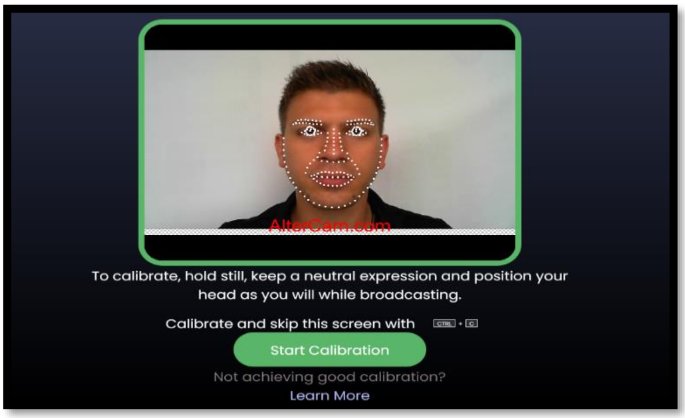
<u>Step 1:</u> Each test cycle was a screen recorded video of the avatar as it imitated the demo video source. These tests differed as each time there was ONE (1) facial feature changed, with the rest of the features being the above detailed "Default".



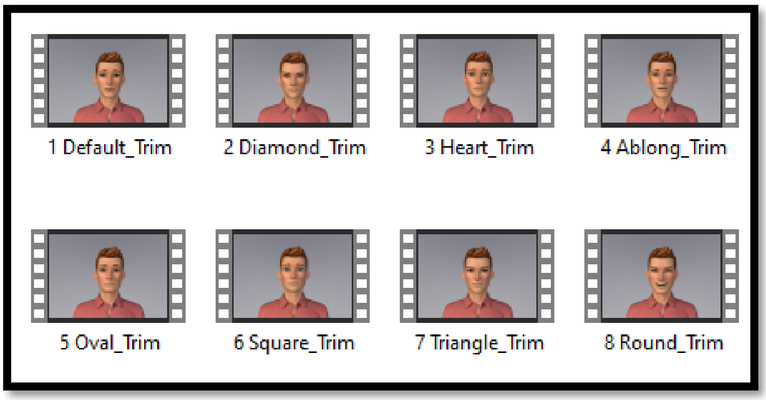
<u>Step 2:</u> At the start of each test cycle, the demo video was reset to play from the beginning. This showed the neutral state as seen below.



<u>Step 3:</u> While showing the neutral state, the "Calibrate" option was used. This ensured Animaze had calibrated to the demo video during each cycle.



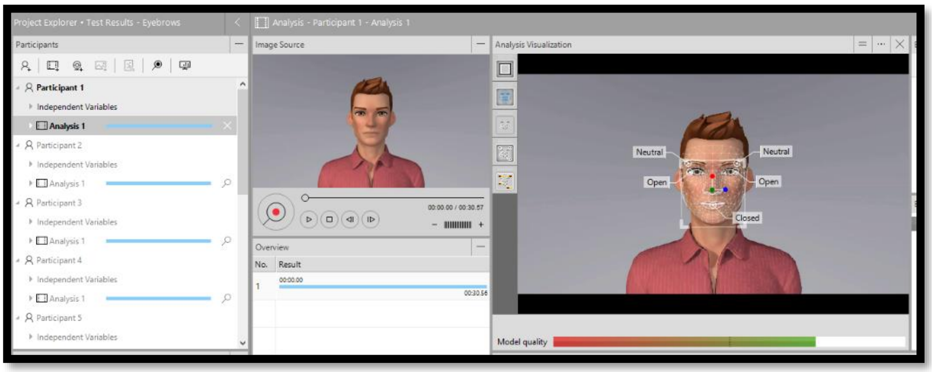
<u>Step 4:</u> Once the source video was played fully, the screen recording of the avatar was exported as an MP4 file and placed into a folder. The naming convention used was 'PX: *Feature Name*' where X was the participant number assigned in ascending order when each new feature was added, and *Feature Name* was the Animaze description of what each feature was called.



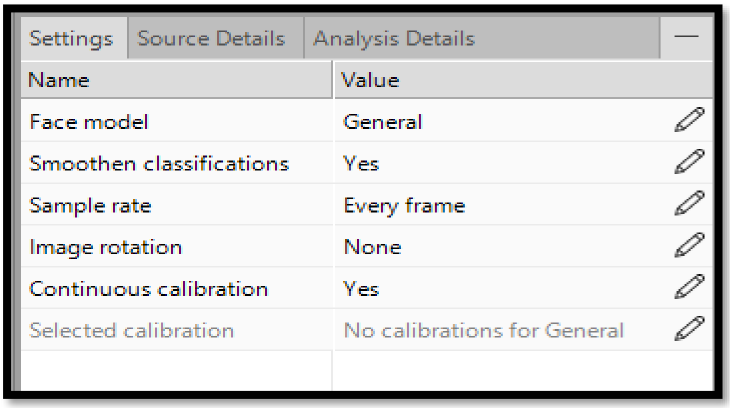
<u>Step 5:</u> Once each facial feature in one group was tested, the avatar was reset to the default state before the next set of features were used. The cycle was repeated until all available facial features were used and exported as an MP4 file.

4.4.3 Testing: FaceReader

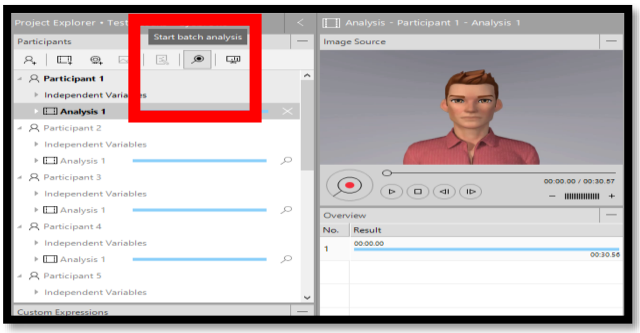
<u>Step 1:</u> The participant videos were categorized according to facial features (Face Shape, Eyebrow, Eyes, Ears, Nose, Mouth) before being transferred into the FaceReader program.



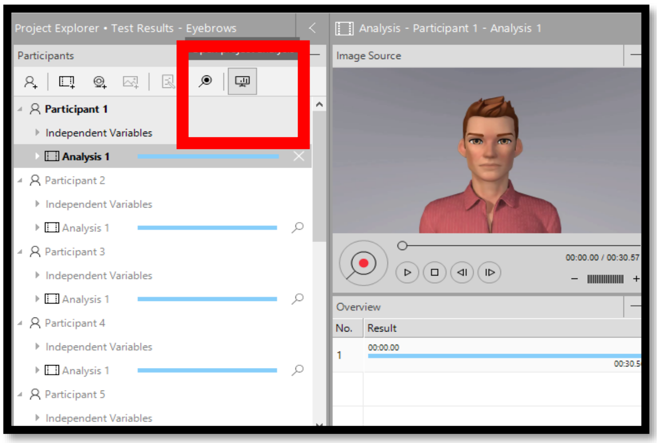
<u>Step 2:</u> Each video was introduced to *FaceReader* as a “Participant” with an Analysis aspect. These participants were given the “Male” gender as an independent variable. The “Continuous Calibration” option was the only setting changed from the *FaceReader* default settings, whereas the other options in *FaceReader* were left as their default options. The preference for continuous calibration was explained above in **Section Four.**



<u>Step 3:</u> Once all participants are introduced and set up, the ‘Batch Analysis’ option was used as this would analyse an entire participant group at once.



<u>Step 4:</u> After analysis was completed, the “Project Analysis” option in *FaceReader* was used to export the results of the *FaceReader* analysis.



4.5 Analysis Measurement

Figure 6 displays the general information that was used to rationalise the results presented. This figure is used as a baseline for the presented research analysis and why facial features may or may not have had an impact on the detected average intensities.

EXPRESSION	AREA OF FOCUS	**BROW AREA**	EYES	**NASAL**	MOUTH
HAPPY	Lower Face	**No distinct brow movement**	- Lower Eyelid may be raised - Crows feet on corner of eyes	**Wrinkles/contraction around nasal area from raised cheek**	-Corner of lips drawn up and back -Mouth more open with more intense emotion. -Teeth may be exposed.
SAD	Entire Face, Eyebrow focus (mouth may be used to cover sadness resulting in happy-sad)	**-Eyebrows are drawn closer together -Inner eyebrows move upwards in triangular fashion**	-Inner corner of upper eyelid is raised -Lower eyelid can be raised in more intense examples	**Little change, slight contraction between eyes**	Corners of lips are drawn down, can be trembling
ANGRY	Entire face, with intensity showing in the mouth being opened wider	**Eyebrows lowered and drawn together**	-Upper and lower eyelids are tense -Upper lid may be lowered due to brow -Lower lid may or may not be raised, iris is focal point	**More intense anger results in more wrinkles and contraction of nasal area**	-Lips are either firmly pressed together, corners straight down -Or can be opened in a square shape (as if shouting)
SURPRISED	Entire face, with intensity found in the lower jaw dropping further	**Eyebrows raised, curving, and raising higher on the face**	-Eyes open wide, white of the eyes seen -Lower eyelid relaxed and upper eyelid raised	**Relaxed nasal area, muscles move outwards**	-Jaw drops, teeth and lips are parted -Opened mouth is relaxed rather than tense
SCARED	Entire face but intensity is shown in the mouth and rising of upper eyelid	**-Eyebrows raised and drawn together -Inner eyebrows closer when compared with surprise**	-Eyes are opened and tensed -Upper eyelid raised, lower eyelid raised and tense	**Relaxed mostly but more intense fear causes stretching around lower nasal.**	-Mouth open, lips are tensed or stretched -Mouth may seem drawn back and tight when compared with surprise
DISGUSTED	Middle area, focus on nasal	**Brow is lowered**	-Lowered brow causes lowered upper eyelid -Wrinkles will appear below lower eyelid	**-Area is wrinkled as face contracts -Cheeks are raised high**	-Lower lip varies as it can be raised and pushed into upper lip -or lowered and protrudes outwards

Figure 6 Summary of the Traits of Emotional Expression

4.6 Conclusion

This section presented the materials used and the procedures that were followed for the presented research. The programs used for this study were *FaceReader*, a computer vision program that presents facial recognition abilities, and *Animaze*, a virtual avatar library. The approaches to this study were laid out in this section, followed by a step-by-step guideline for how this study could be recreated for future research. This was followed by a table of information which presented a summary of information for how the facial expressions of the virtual avatars would be interpreted by a human participant. The following section will be presenting the results of the experiment laid out in this section, which will also include a discussion and analysis of these results to best answer the presented research questions.

Section 5: Results and Analysis

5.1 Introduction

The findings presented below have been provided through the use of 78 different screen recorded videos of virtual avatars, which will be referred to as 'participants' throughout this section. These videos were organised into 8 different participant groups to better organise the research findings and to assist the answering of the research question by providing a scoped in focus on the results provided by *FaceReader*. The participant groups were divided up as follows:

1. 8 Face Shape participants
2. 11 Eyebrow participants
3. 11 Eye participants
4. 12 Ear participants
5. 12 Nose participants
6. 12 Mouth participants
7. 6 participants that were modelled to INCREASE one of the six basic emotions
8. 6 participants that were modelled to DECREASE one of the six basic emotions

These participants were analysed through *FaceReader* and compiled into a data sheet to assist the visualisation and analysis of the results. The participants were separated beginning with the six facial shape groups, followed by the two groups of six virtual avatars, where group 7's avatars were designed with facial features that exhibited the highest average intensity of each emotional expression and group 8's avatars were designed with facial features that exhibited the lowest average intensity of each emotional expression. These groups were designed to test the viability of creating an intentional personal bias within the virtual avatars to either increase or decrease the six basic emotional expressions. The original readings provided through *FaceReader* measured the average intensity of the emotions on a scale of 0-1 throughout the 30 second video, as outlined in **Section 4.** As several readings were in decimals as small as 0.0001, the recorded values were multiplied by 100 to change the average intensity scale from 0-1 to 1-100. The reasoning for this scale change was to provide numbers that are easily understandable, as the original readings were so small, and to make the glance value of the resulting graphs easier to read.

The results in the presented research were separated into two sections: **Section 5.1** is the 'Facial Feature' focused test, and **Section 5.2** is the 'Emotional Expression' focused test. The 'Facial Feature' focused test will be analysing the results as they are received from the original testing groups (**Table 1** would cover the 'Face Shape' facial feature, **Table 2** would cover the 'Eyebrow' facial feature group, and so on) while the 'Emotional Expression' focused test would only focus on a single emotional expression at a time (Graph 1 would show every 'Happy' average intensity from the results in Table 1-6, Graph 2 would show only 'Sad' average intensities, and so on).

These two sections were used for the presentation of the results as the focus of the presented data would be able to answer different research questions more effectively. To clarify, the results found in the two sections are the same and were unaltered, the difference is found in how the results are grouped together and what format the results are displayed in. As an example, the 'Facial Feature' testing group results would be more helpful when looking to answer, *"How would the shape of the nose affect FaceReader's ability to detect 'Happy' as an emotional expression?"* as the table format clearly indicates the exact facial feature and how they affect the emotional expressions. Additionally, the 'Emotional Expression' focused group would portray graphs that would be more helpful in answering the question of, *"Which specific facial feature had the Highest Average Intensity reading for 'Happy' across all participants?"* as the graphs combine all the of the readings for that emotional expression into one graph. The use and aim of the two sections will be clarified in **Section 5.2** onwards.

5.2 Results: Feature Focus

5.2.1 Introduction

Section Three discussed how the Animaze virtual avatar library provides the user with a large amount of control over the editing of facial features for its virtual avatars. This section will be presenting 66 facial features available in *Animaze* by discussing their effect on *FaceReader's* facial expression recognition. These facial features will be split into six groups: Face Shape, Eyebrows, Eyes, Ears, Nose, and Mouth, where each group will discuss the facial feature that was altered, how this alteration affected *FaceReader's* expression recognition, and thoughts on the reasoning for these interactions. The overall aim of this section is to identify the emotional expressions that each facial feature group is most corelated with. This would be done by finding within these testing groups, which emotional expressions have a higher difference in presence and intensity with varying facial features.

Additionally, this section will end off with an analysis of the ranges found in each testing group to further identify the facial features that had a greater effect on the expression recognition to answer how *FaceReader* may be affected by different facial features on a virtual avatar. What should be clarified again is that intensity is different from appearance. The presence of the emotional expression is if *FaceReader* detects that emotional expression being shown by the virtual avatar, while intensity is how vivid the presence of the emotional expression is. The readings presented below are the intensity averages taken from *FaceReader* as they provide a more detailed look at how these facial features affect expression recognition, rather than the focus being on the accuracy of *FaceReader's* expression recognition.

5.2.2 Analysis 1: Participant groups were facial features

The results presented in the presented tables would be read as follows:

- Each **ROW** is a virtual avatar with a different facial feature, with the names of the features indicated in the left most column.
- Each **COLUMN** indicates one of the six basic emotion expressions (from left to right: Happy, Sad, Angry, Surprised, Scared, and Disgusted) with the average intensities being anywhere between 0-100.
- The **GREEN** blocks indicate the *Highest Average Intensity* (*HAI*) of each expression. This indicates which facial feature within that testing group produced the *HAI* of each emotional expression.
- The **YELLOW** blocks indicate the *Lowest Average Intensity* (*LAI*) of each expression. This indicates which facial feature within that testing group produced the *LAI* of each emotional expression.
- The emotional expression of 'Neutral' is available as a reading option and it is very prevalent in this analysis, but 'neutral' is not one of the six basic emotions that is being tested for and as such was removed.
- The second to last rows consist of the overall *mean* (average) for the detected expression intensities, which was calculated by dividing the sum of the average intensities in one column and dividing them by the amount of participants.
- The last row presents the *range* of the expression intensities, which was calculated by subtracting the LAI (yellow value) from the HAI (green value) in each column.

This section will provide each table in full, followed by a summary of the overall average intensities, the Highest Average Intensities, the Lowest Average Intensities, and the ranges.

These points will be highlighted as they are the readings that will be used in the analysis sections, which will follow each table, and these readings will be referred to in the discussion section. The analysis of the tables will be done through using the analysis measurement chart provided at the end of **Section Four.** This will be done through analysing which facial feature provided the Highest Average Intensities to determine how the Highest Average Intensities would have come about, with an analysis of the Lowest Average Intensities being similar as I will theorise how that facial feature produced a lower reading. When there is a noteworthy outlier, the mean and range readings will be mentioned but they will mainly be used in the 'Discussion' sub-section as the value in these readings are found in how they compare to the averages and ranges of the other facial features.

It is important to note that the input for each of these participants is always the same (the 'Demo Video'), which indicates that the way the features alter the virtual avatars face and *FaceReader's* perception of the face are what changes the expression recognition readings. The general ranges of the tables below are **between 0 and 26 on a scale of 0-100,** as the highest reading is 25.56 (P15 in 'Surprised') and the lowest reading is 0.08 (P18 in 'Scared'). The analysis of these tables will be on identifying the outliers of average intensities that each facial feature group has produced, and to discuss what the characteristics of those facial features are. As an example, if the 'Nose' group produced notably higher average intensities of the 'Sad' emotional expression, the intent is to isolate the features that produced these readings and discuss what their characteristics are. This would assist the answering of the research question by providing insight into which facial features are able to alter *FaceReader's* analysis and what the reasoning may be.

5.2.3 Face Shape Testing Group

Table 1 Face Shape

Test 1: Face Shape	Happy	Sad	Angry	Surprised	Scared	Disgusted
P1: Default	3,79	1,32	13,19	21,85	0,16	0,95
P2: Diamond	4,18	0,81	11,81	21,99	0,22	0,86
P3: Heart	4,15	0,76	12,57	23,55	0,18	1,6
P4: Oblong	4,4	0,86	11,45	17,82	0,35	1,54
P5: Oval	3,45	1,18	12,02	16,57	0,47	1,53
P6: Square	2,54	0,81	9,65	19,57	0,27	2
P7: Triangle	3,76	0,67	7,84	18,81	0,32	2,01
P8: Round	3,57	1,49	9,63	17,07	0,48	1,34
Mean	3,73	0,99	11,02	19,65	0,31	1,48
Range	1,83	0,82	5,35	6,98	0,32	1,18

The average intensity of the expression “Happy” being detected was found to be 3.73, with the HAI being 4.4, presented in the “Oblong” participant, and the LAI being 2.54, presented in the “Square” participant. A range of 1.83 was found in the “Happy” expression intensity when altering the face shape.

The average intensity of the expression “Sad” being detected was found to be 1.32, with the HAI being 1.49, presented in the “Round” participant, and the LAI being 0.67, presented in the “Triangle” participant. A range of 0.82 was found in the “Sad” expression intensity when altering the face shape.

The average intensity of the expression “Angry” being detected was found to be 11.02, with the HAI being 13.19, presented in the “Default” participant, and the LAI being 7.84, presented in the “Triangle” participant. A range of 5.35 was found in the “Angry” expression intensity when altering the face shape.

The average intensity of the expression “Surprised” being detected was found to be 19.65, with the HAI being 23.55, presented in the “Heart” participant, and the LAI being 16.57, presented in the “Oval” participant. A range of 6.98 was found in the “Surprised” expression intensity when altering the face shape.

The average intensity of the expression “Scared” being detected was found to be 0.31, with the HAI being 0.48, presented in the “Round” participant, and the LAI being 0.16, presented in the “Default” participant. A range of 0.32 was found in the “Scared” expression intensity when altering the face shape.

The average intensity of the expression “Disgusted” being detected was found to be 1.48, with the HAI being 2.01, presented in the “Triangle” participant, and the LAI being 2.54, presented in the “Diamond” participant. A range of 1.83 was found in the “Disgusted” expression intensity when altering the face shape.

5.2.3.1 Table 1: Face Shape – Analysis

As clarified in **Section 4**, computer vision machines do not specifically look at the outlining of the face to determine the intensity of any emotional expressions. With that being said, the shape of the face was found to change the size of the surface area around the mouth and cheeks, which are important areas of analysis for several emotional expressions.

The face shapes with sharper jaw lines (P4, P3, P2) which reduced the surface area of the cheeks were found to produce higher ‘Happy’ intensities. This is due to *FaceReader* interpreting the mouth as opening wider and producing a bigger smile, when in fact there is just less surface area on the cheeks making an illusion of a wider smile. This effect worked in a similar fashion for the 'Sad' intensities, as the face shapes that were rounder would produce a higher 'Sad' Average Intensity. This was found to be due to the rounder facial features (P8, P5) producing more surface area for the mouth to cover, making a wide smile appear to be less full when analysed by FaceReader.

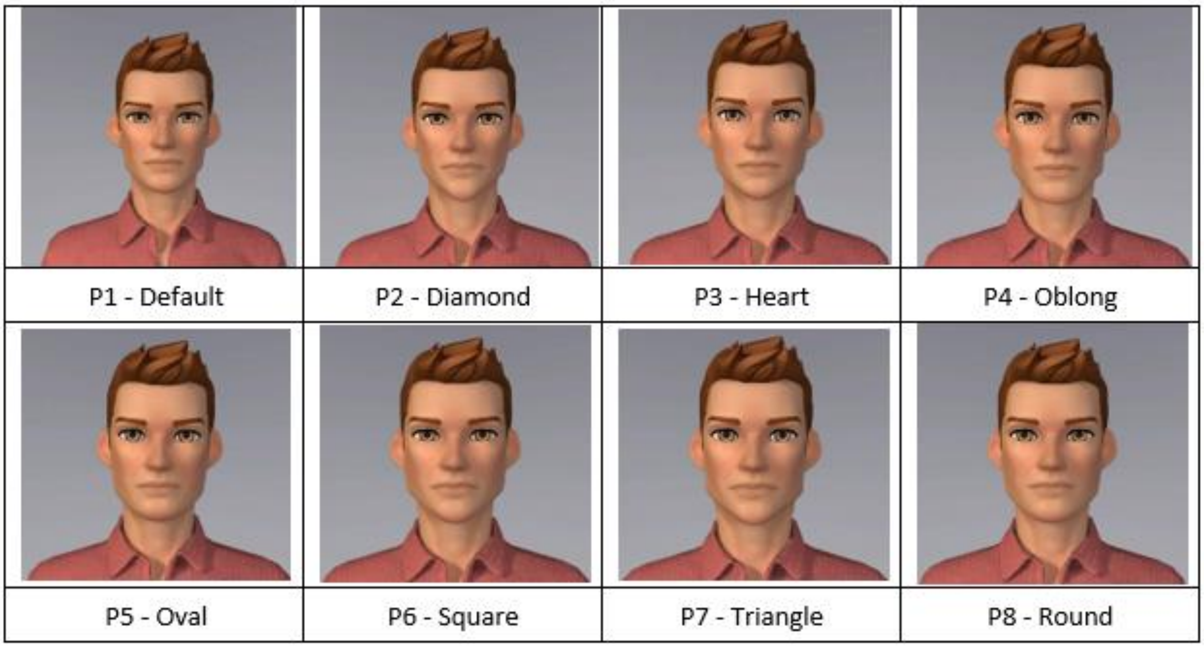


Figure 7 The 'Face Shape' participants.

The 'Angry' Average Intensities are harder to interpret as anger an emotion that is focused on the eyebrows, with the mouth area playing a larger role in differentiating the form of anger (verbal Vs. Physical) but it is noted that anger can be expressed through how deeply the sides of the mouth drop. In the face shapes that raise the chin higher (P1, P3, P5) the mouth would be interpreted as moving lower when really there is just less surface area to cover. The Average Intensities of 'Surprise' can be found to be higher in Face Shapes that have a slightly higher and more pointed chin (P2, P3). This would be due to surprise being read through how far the jaw drops and widens, making FaceReader interpret a jaw that starts from a higher point as being more surprised when compared with a chin that starts from a lower point due to its wider shape (P5, P4).

'Scared' Average Intensities were found to be higher in Face Shapes that were rounder (P8, P5) and lower in sharper and smaller face shapes (P1, P3). This was interpreted as being due to FaceReader interpreting the wider mouth areas as the virtual avatar stretching the mouth wider, which results in a higher intensity of 'Scared' being found. Disgust was found to have higher Average Intensities in face shapes that had sharp but wide jaw lines (P6, P7) and lower Average

Intensities in sharp but thin jaw lines (P1, P2). This would be due to FaceReader interpreting the wider face shapes as having a more contracted mouth, due to the wider surface area around it, while the wider surface area of the cheeks would be interpreted as having a less contracted mouth (resulting in less 'Disgust' being found).

5.2.3.2 Table 1 - Summary

The results found in the 'Face Shape' Average Intensities showed that the importance of this facial feature lies less in the shape of the face and more in how the face shape would increase the surface area of the cheeks and the other areas around the mouth. It is possible to create a personal bias for certain emotional expressions that rely on the amount of movement of the mouth as the larger surface areas can change FaceReader's interpretation of how far the areas around the mouth have moved from their default position. Face shapes with sharper features tend to lean toward 'Happy' and 'Surprise' as their intensities rely on how wide the mouth is open. Rounder face shapes would lean toward 'Sad' and 'Scared' due to these expressions relying more on how the mouth lowers and contracts to take up less surface area. 'Anger' and 'Disgust' would have a general mix of a wider face but with a shaper jaw line and the intensity of these emotional expressions change with the shape of the mouth rather than how wide it is opened.

5.2.4 Eyebrows Testing Group

Table 2 Eyebrows

Test 2: Eyebrows	Happy	Sad	Angry	Surprised	Scared	Disgusted
P9: Default	3,79	1,32	13,19	21,85	0,16	0,95
P10: Eyebrow #1	4,68	0,18	7,75	25,53	0,09	1
P11: Eyebrow #2	4,04	0,29	6,15	20,54	0,18	1,77
P12: Eyebrow #3	3,25	0,94	10,42	18,06	0,23	1,3
P13: Eyebrow #4	4,08	0,31	7,62	23,86	0,1	1,12
P14: Eyebrow #5	4,29	0,25	6,08	24,23	0,16	1,73
P15: Eyebrow #6	3,49	0,81	11,47	17,73	0,28	1,85
P16: Eyebrow #7	3,56	0,38	12,28	22,28	0,08	1,78
P17: Eyebrow #8	4,21	0,49	8,8	22,15	0,13	2,14
P18: Eyebrow #9	3,53	0,36	6,09	22,43	0,08	1,77
P19: Unibrow	3,82	0,69	3,6	25,56	0,2	0,76
Mean	3,89	0,55	8,50	22,20	0,15	1,47
Range	1,43	1,14	9,59	7,83	0,2	1,38

The average intensity of the expression "Happy" being detected was found to be 3.89, with the HAI being 4.68, presented in the "Eyebrow #1" participant, and the LAI being 3.25, presented in the "Eyebrow #3" participant. A range of 1.43 was found in the "Happy" expression intensity when altering the eyebrows.

The average intensity of the expression “Sad” being detected was found to be 0.55, with the HAI being 1.32, presented in the “Default” participant, and the LAI being 0.18, presented in the “Eyebrow #1” participant. A range of 1.14 was found in the “Sad” expression intensity when altering the eyebrows.

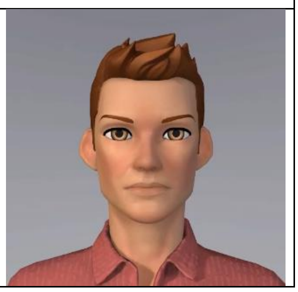
The average intensity of the expression “Angry” being detected was found to be 8.50, with the HAI being 13.19, presented in the “Default” participant, and the LAI being 3.6, presented in the “Unibrow” participant. A range of 9.59 was found in the “Angry” expression when altering the eyebrows.

The average intensity of the expression “Surprised” being detected was found to be 22.20, with the HAI being 25.56, presented in the “Unibrow” participant, and the LAI being 17.73, presented in the “Eyebrow #6” participant. A range of 7.83 was found in the “Surprised” expression intensity when altering the eyebrows.

The average intensity of the expression “Scared” being detected was found to be 0.15, with the HAI being 0.28, presented in the “Eyebrow #6” participant, and the LAI being 0.08, presented in the “Eyebrow #9” participant. A range of 0.2 was found in the “Scared” expression intensity when altering the eyebrows.

The average intensity of the expression “Disgusted” being detected was found to be 1.47, with the HAI being 2.14, presented in the “Eyebrow #8” participant, and the LAI being 0.76, presented in the “Unibrow” participant. A range of 1.38 was found in the “Disgusted” expression intensity when altering the eyebrows.

5.2.4.1 Table 2: Eyebrows– Analysis

			
P9 - Default	P10 – Eyebrow #1	P11 – Eyebrow #2	P12 – Eyebrow #3
			

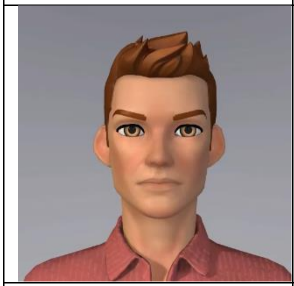
P13 – Eyebrow #4	P14 – Eyebrow #5	P15 – Eyebrow #6	P16 – Eyebrow #7
			
P17 – Eyebrow #8	P18 – Eyebrow #9	P19 – Unibrow	

Figure 8 The 'Eyebrows' participants.

Section 4 explained the importance that the eyebrows have in the expression of several of the six basic emotions. The HAI of 'Happy' was found in eyebrows that were arc shaped and were more rounded (P10, P14, P17) whereas the LAIs were found in eyebrows that were straighter and angled inwards (P12, P16). While 'Happy' is not said to be determined by the eyebrows, the straighter and angled eyebrows would be interpreted as less happy as *FaceReader* would bias these eyebrows to being 'Angry'. The HAI for 'Happy' and the LAI for 'Sad' (P10) indicated that the eyebrows are not leaning towards being happy, rather they are leaning away from sadness and anger. Additionally, the HAI for 'Sad' was found in the default avatar (P9) as the default eyebrows are angled to form an almost straight line from the end of the one eyebrow to the end of the other. This would cause *FaceReader* to interpret them as sadder when compared with the other eyebrow options as the other eyebrows begin in an angled inward position.

The default eyebrows (P9) were also found to have the HAI of 'Anger' as these eyebrows begin from a more even angle. This would make *FaceReader* interpret the inwards angling of the default eyebrows to be angrier, while the LAI found in the already angled eyebrows (P18, P19) would be due to the calibration accounting for the inward angles already making the appearance of anger less prevalent. 'Surprise' is also affected by the curved shape of the eyebrows (P19, P14) as the eyebrows that have a default position that arcs upward were found to lean toward expressing surprise, while the eyebrows that are straight (P15, P12) are less likely to be interpreted as expressing surprise. This bias is due to the intensity of 'Surprise' being found in how raised and how curved the eyebrows are, meaning straighter eyebrows would produce a lower intensity of that emotional expression.

"Scared" had an inverse effect as the HAIs were found in the straighter eyebrows (P12, P15) and the LAIs were found in slightly more curved eyebrows (P10, P18). This was noted to be due to 'Surprise' and 'Scared' being very similar, but 'Scared' eyebrows tend to raise without curving, and they move in closer together. The straighter eyebrows exhibit the inward motion easily as they do not curve inward and away from one another (as seen in 'Surprise'). 'Disgust' in the eyebrows was said to be found in how low the eyebrows are and how they would contract inwards, which is seen in FaceReader interpreting the eyebrows that are lowered inward more (P17) as having the HAI. In comparison, the eyebrows that are straighter or ones that have uneven angles (P9, P19) would be interpreted as exhibiting less 'Disgust' as these eyebrows have less of an ability to move downward and contract inwards.

5.2.4.2 Table 2 - Summary

These results provide evidence towards the importance of the eyebrows when portraying several of the basic emotions. The angle of the eyebrows was found to be the most important aspect, as those with a higher arc (rounded) would be bias towards being seen as expressing 'Happy' and 'Surprise', while the flat (straight) eyebrows would lean towards 'Sad' and 'Angry'. The starting angle of the eyebrows were also found to change FaceReader's interpretation of expressions as the eyebrows which begin from a more inward angle lean towards 'Disgust' and 'Anger' while those with a more outward angle lean towards 'Scared' and 'Sad'.

5.2.5 Eyes Testing Group

The average intensity of the expression "Happy" being detected was found to be 3.78, with the HAI being 4.34, presented in the "Big Upturned" participant, and the LAI being 3.2, presented in the "Upturned" participant. A range of 1.34 was found in the "Happy" expression intensity when altering the eyes.

The average intensity of the expression "Sad" being detected was found to be 1.75, with the HAI being 6.38, presented in the "Small Monolid" participant, and the LAI being 0.54, presented in the "Big Upturned" participant. A range of 5.84 was found in the "Sad" expression intensity when altering the eyes.

The average intensity of the expression "Angry" being detected was found to be 10.48, with the HAI being 13.19, presented in the "Default" participant, and the LAI being 5.34, presented in the "Small Monolid" participant. A range of 7.85 was found in the "Angry" expression when altering the eyes.

Table 3 Eyes

Test 3: Eyes	Happy	Sad	Angry	Surprised	Scared	Disgusted
P20: Default	3,79	1,32	13,19	21,85	0,16	0,95
P21: Almond	4,33	0,98	12,28	16,58	0,5	0,71
P22: Hooded	3,41	1,1	9,15	16,24	0,22	0,76
P23: Monolid	3,34	2,02	7,93	14,73	0,15	1,67
P24: Downturned	3,82	1,14	11,85	17,3	0,24	1,13
P25: Upturned	3,2	1,42	12,33	14,27	0,47	0,96
P26: Small Monolid	4,19	6,38	5,34	9,89	0,32	1,16
P27: Big Monolid	3,66	0,56	11,42	18,17	0,09	1,81
P28: Small Upturn	3,41	2,35	8,87	12,04	0,35	0,74
P29: Big Upturn	4,34	0,54	12,34	19,1	0,19	1,02
P30: Round	4,12	1,45	10,58	15,05	0,41	0,42
Mean	3,78	1,75	10,48	15,93	0,28	1,03
Range	1,34	5,84	7,85	11,96	0,41	1,39

The average intensity of the expression “Surprised” being detected was found to be 15.93, with the HAI being 21.85, presented in the “Default” participant, and the LAI being 9.89, presented in the “Small Monolid” participant. A range of 11.96 was found in the “Surprised” expression intensity when altering the eyes.

The average intensity of the expression “Scared” being detected was found to be 0.28, with the HAI being 0.5, presented in the “Almond” participant, and the LAI being 0.09, presented in the “Big Monolid” participant. A range of 0.41 was found in the “Scared” expression intensity when altering the eyes.

The average intensity of the expression “Disgusted” being detected was found to be 1.03, with the HAI being 1.81, presented in the “Big Monolid” participant, and the LAI being 0.42, presented in the “Round” participant. A range of 1.39 was found in the “Disgusted” expression intensity when altering the eyes.

5.2.5.1 Table 3: Eyes– Analysis

The focus of the eyes when interpreting emotional expression are mainly found in how wide the eyelids have opened, the visibility of the iris, and how the brows have frowned to cover the eyes. This creates a challenge as Animaze does have advanced movements with the eyes and eyelids, the ability for the program to read and imitate tiny details in the eyelids may cause problems in how emotional expressions are interpreted.

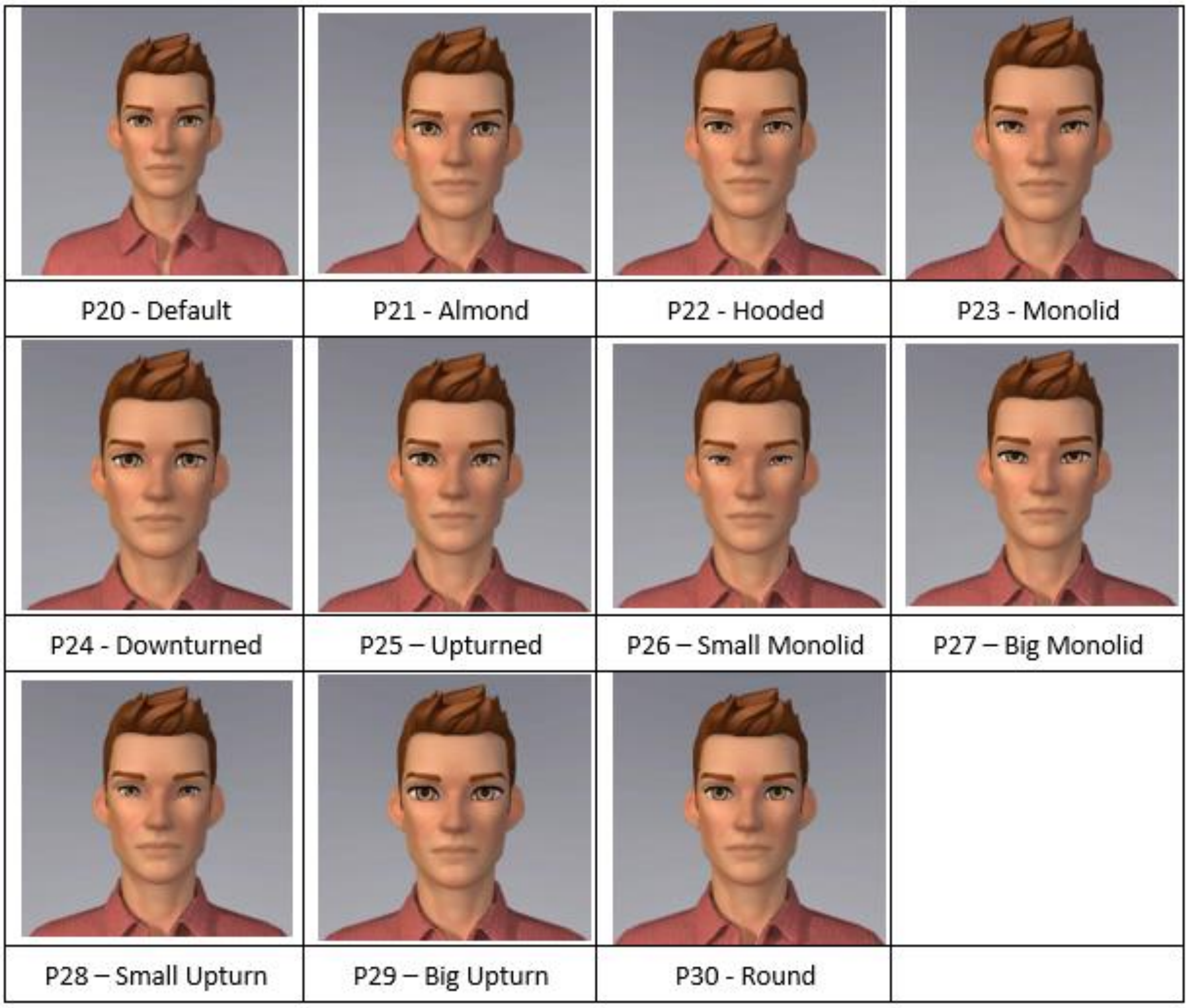


Figure 9 The 'Eye' participants.

"Happy" has it's HAI in the eyes that are larger (P29, P21) and are angled in a straight line (not leaning inwards or outwards) while the LAI of 'Happy' is seen in eyes that are smaller (P23, P25). This would be due to the larger eye shapes being interpreted as having higher raised upper eyelids due to the larger size of the eye making movements more visible. The smaller eyes also have a harder time portraying the 'crows-feet' aspect as they do not extend all the way out to the side of the head. In a similar fashion, 'Sad' is expressed with a very high Average Intensity in the smaller eyes that also have dropping upper eyelids (P23, P26) while the LAI of 'Sad' is expressed through all of the larger eyes. This again would be due to the smaller eye shapes being interpreted as having lowered eyelids, when the reality is the eyelids are just smaller and would be less noticeable.

'Anger' and 'Surprised' share a HAI (P20) and a LAI (P26), which is interesting as they do not share similar details that are looked for when interpreting these expressions. 'Anger' would look for a more squinted eye shape, whereas 'Surprised' would be based on how wide the eye is opened. What this is interpreted as is that the default eye shape has the highest control over the eyelids, and FaceReader is able to easily interpret the position of the eyelids when these emotional expressions are being presented. Additionally, the shared LAI (P26) is the smallest eye shape and would have the least amount of visible control over the eyelids. This reduces the visibility of the iris, and important factor for both ‘Anger’ and ‘Surprised’.

The rounder eyes (P21, P25, P30) were found to lean towards portraying ‘Scared’, as the eyes that are shaped to be sharper (P23, P27) are found to lean away from looking ‘Scared’. This difference was attributed to the rounder eyes being interpreted by FaceReader as being naturally ‘Surprised’ as more of the iris is visible, whereas the sharper eyes would not be as wide in their openings. The inverse was found for ‘Disgusted’ as the sharper eye shapes (P23, P27) would lean towards portraying disgust due to the easier wrinkling under the eyes, and the upper eyelid would tend to be lowered already. Additionally, the eye shapes that are rounder and are angled with the outer areas leaning downwards (P29, P30) would lean away from ‘Disgusted’ as FaceReader would detect less wrinkling under the eyes and the upper eyelid would not be lowered as much as the other eye shapes.

5.2.5.2 Table 3 - Summary

The importance of the eye shape was found in the size of the eye, the angle of the eye shape, and the positioning of the upper eyelid. A larger eye would easily portray ‘Happy’, while large round eyes would lean towards ‘Scared’. Smaller eyes would tend to portray ‘Sad’ while the smaller and angled eye shapes would exhibit an increase in Disgust’. ‘Anger’ and ‘Surprise’ had the same eye shape as their HAI and LAI readings matched one another, even though they search for different movements and details. What this found was that the eye shape would heavily impact Animaze’s ability to control the eyelids, and *FaceReader’s* ability to read and interpret these eyelid movements. The visibility of the iris was almost always important (aside from in ‘Sad’) as how visible the iris is would be interpreted as how the eyelids were moved.

5.2.6 Ears Testing Group

Table 4 Ears

Test 4: Ears	Happy	Sad	Angry	Surprised	Scared	Disgusted
P31: Default	3,79	1,32	13,19	21,85	0,16	0,95
P32: Elf Med	3,61	1,18	8,93	20,4	0,17	1,24
P33: Scale Med	4,54	1,37	7,66	19,62	0,12	1,79
P34: Front Med	5,07	1,43	8,11	20,12	0,17	1,33
P35: Default Small	3,93	1,88	8,81	20,11	0,22	2,34
P36: Elf Small	4,55	1,97	7,7	18,9	0,22	1,8
P37: Scale Small	4,16	1,04	7,31	18,61	0,24	2,03
P38: Front Small	3,88	0,85	10,79	21,21	0,19	1,98
P39: Default XL	2,77	0,78	9,29	20,42	0,12	1,86
P40: Elf XL	2,61	0,81	10,95	20,32	0,09	1,3
P41: Scale XL	3,25	0,62	9,78	18,93	0,19	1,54
P42: Front XL	3,79	1	7,59	22,06	0,12	1,55
Mean	**3,83**	**1,19**	**9,18**	**20,21**	**0,17**	**1,64**
Range	2,46	1,35	5,88	3,45	0,15	1,39

The average intensity of the expression “Happy” being detected was found to be 3.83, with the HAI being 5.07, presented in the “Front Facing - Med” participant, and the LAI being 2.61, presented in the “Elf - XL” participant. A range of 2.46 was found in the “Happy” expression intensity when altering the ears.

The average intensity of the expression “Sad” being detected was found to be 1.19, with the HAI being 1.97, presented in the “Elf - Small” participant, and the LAI being 0.62, presented in the “Scale - XL” participant. A range of 1.35 was found in the “Sad” expression intensity when altering the ears.

The average intensity of the expression “Angry” being detected was found to be 9.18, with the HAI being 13.19, presented in the “Default” participant, and the LAI being 7.31, presented in the “Scale - Small” participant. A range of 5.88 was found in the “Angry” expression when altering the ears.

The average intensity of the expression “Surprised” being detected was found to be 20.21, with the HAI being 22.06, presented in the “Front Facing - XL” participant, and the LAI being 18.61, presented in the “Scale - Small” participant. A range of 3.45 was found in the “Surprised” expression intensity when altering the ears.

The average intensity of the expression “Scared” being detected was found to be 0.17, with the HAI being 0.24, presented in the “Scale - Small” participant, and the LAI being 0.09, presented in the “Elf - XL” participant. A range of 0.15 was found in the “Scared” expression intensity when altering the ears.

The average intensity of the expression “Disgusted” being detected was found to be 1.64, with the HAI being 2.34, presented in the “Default - Small” participant, and the LAI being 0.95, presented in the “Default” participant. A range of 1.39 was found in the “Disgusted” expression intensity when altering the ears.

5.2.6.1 Table 4: Ears– Analysis

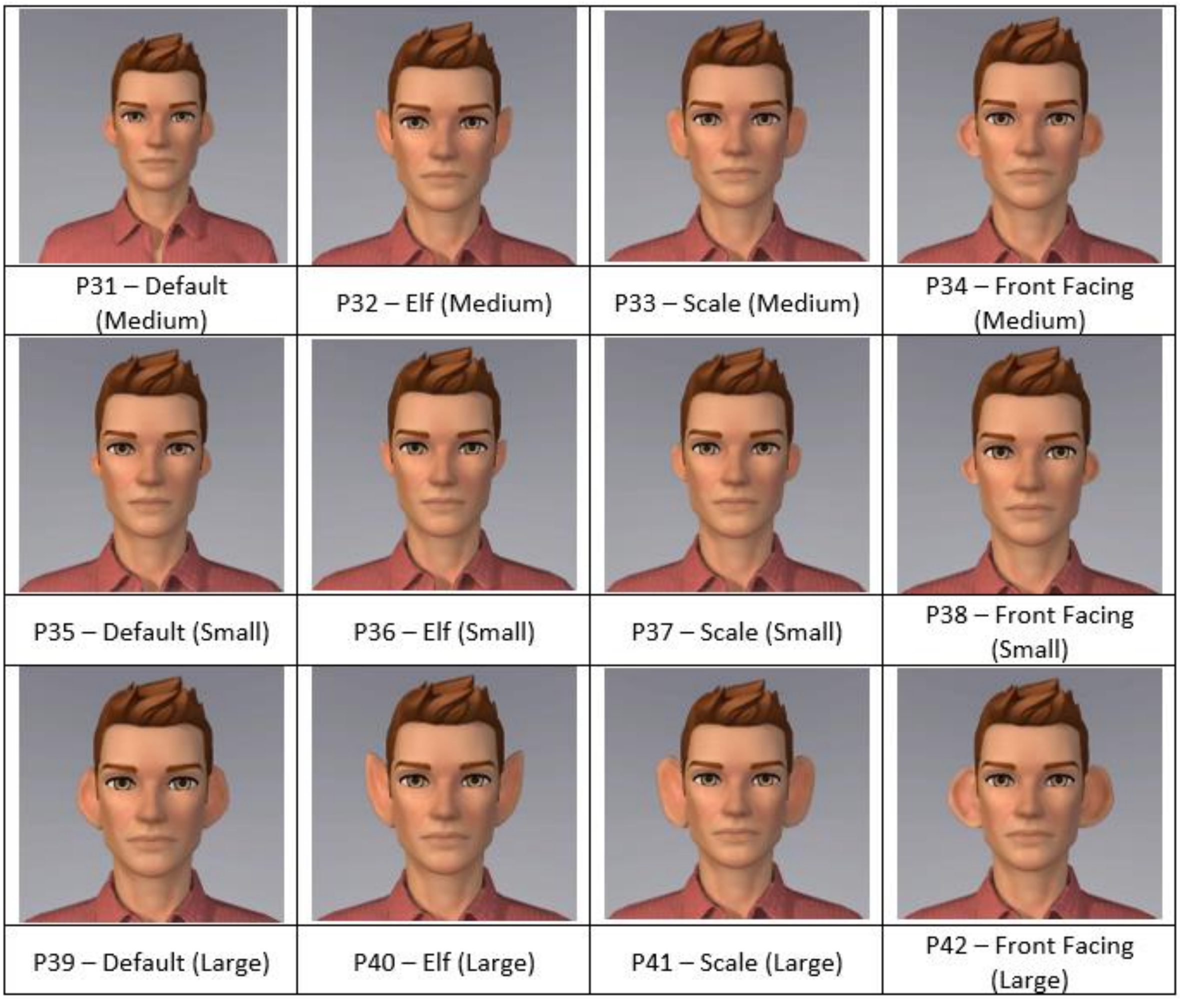


Figure 10 The ‘Ear’ participants.

Similar to the ‘Face Shape’ facial feature, ‘Ears’ are not established as an area of focus when interpreting emotional expressions. This would be due to there being no intentional movements of the ears in human emotional expressions as we lack the ability to control the movements of our ears in the same way animals, such as dogs, are able to. As such, the analysis of these results will be done through discussing how the shape and sizes of the ears would affect the areas around the cheeks and eyes of the face to rationalise how different Average Intensities were produced from a change in the ear facial feature.

‘Happy’ was found to be more prevalent when the medium sized, front-facing ears were used (P33, P34) and less prevalent when the larger sized, flatter ear shapes were used (P40, P41). This difference would be due to the front facing ears being more easily identified by FaceReader as an extension of the face and not included in the expression analysis, while the larger flat ears would reduce FaceReader’s ability to differentiate where the outside of the face ends and the ear begins. If the face is viewed as being larger, FaceReader would identify a lower emotional intensity of ‘Happy’ as the identified mouth and eyebrow movements (explained above) would be less noticeable. In a similar sense, ‘Sad’ had a HAI in the smaller ears (P35, P36) and a LAI in the larger ears (P40, P41). This indicates that there is a corelation between the size of the ears in how ‘Sad’ is perceived by FaceReader, possibly due to intensity of the emotional expression being harder to read with larger ears.

The default ears (31) had a noticeably higher Average Intensity of ‘angry’ compared with the other ear choices, with the next highest readings coming from vastly different ear shapes and sizes (P38 & P40). ‘Angry’ also had a variation issue in the LAI readings having few corelations between them. I would also argue that the ears that produced the HAI (P31) and LAI (P37) of ‘Angry’ are extremely similar in their shape and sizes, with the only difference found being that the HAI ear is more rounded, while the LAI is pointed by the earlobe. This would suggest that having a visible earlobe would greatly increase FaceReader’s ability to differentiate the ear from the rest of the face. ‘Surprise’ was found to have an HAI in the front facing ears of all sizes (P34, P38, P42) and a LAI in the scale ears of all sizes (P33, P37, P41). This is interesting as few other times did the size of the ears do not have a drastic effect on the expression intensities, hinting that the shape of the ear is more important when portraying ‘Surprise’.

The smaller ears (P35, P36, P37) produced a HAI of ‘Scared’ when compared with the apparent LAI of this emotional expression found when using the larger ear sizes (P39, P40, P42). This would indicate that the ‘Scared’ emotional expression is analysed more through the size of the ears rather than the shape. The size of the ears had a similar effect on the Average Intensity reading of ‘Disgust’ as well, as the smaller ear sizes (P35, P37, P38) produced a HAI of this emotional expression when compared with the LAI found in the medium sized ears (P31, P33, P34). The difference in Average Intensities of ‘Disgust’ stood out as the other emotional expressions would lean to either the very large or very small sizes (aside from outliers where the medium ears were a HAI), yet ‘Disgust’ was the least prevalent when using an average sized ear size. This could be due to the medium ears being harder to distinguish as disgust, causing a bias toward 'Sad or 'Anger.

5.2.6.2 Table 4 - Summary

In theory, the 'Ears' facial feature should not have any significant effects on the average emotional intensities, yet the presented results inform us that the size and shape of the ear will affect how computer vision programs interpret emotional expressions. Whether or not the ear protrudes out and forwards will affect the average intensity reading of 'Happy' as an ear shape that is flat and closer to the side of the head produces a lower 'Happy' intensity. 'Surprise' was also found to change depending on the shape of the ear, as the same ear shape would, on average, produce the HAI and LAI of 'Surprise' regardless of the size of the ear. The intensity of 'Sad' was found to change depending on the size of the ear, with smaller ears being analysed as being more 'Sad'. The size of the ear was also found to effect the intensity of 'Scared' and 'Disgust' as well, as smaller ears produced a HAI of 'Scared' while the average (medium) sized ears would produce a HAI of 'Disgust". The emotional expression of 'Anger' was the biggest outlier in these results as there were few corelating factors in the ears that led to a HAI and LAI reading. The only noted effect that would cause the expression intensity to increase was if there was a clear distinction between the earlobe and where the ear connects to the side of the face.

5.2.7 Nose Testing Group

Table 5 Nose

Test 5: Nose	Happy	Sad	Angry	Surprised	Scared	Disgusted
P43: Default	3,79	1,32	13,19	21,85	0,16	0,95
P44: Nose #2	3,5	1,82	9,79	16,49	0,4	3,25
P45: Nose #1	4,29	1,85	11,26	19,78	0,36	1,09
P46: Nose #3	4,06	2,75	10,83	20,38	0,42	2,19
P47: Nose #4	3,91	1,51	11,13	18,85	0,56	2,47
P48: Nose #5	2,97	1,46	10,61	16,19	0,99	8,97
P49: Nose #6	3,2	0,6	10,58	17,75	0,36	7,39
P50: Nose #7	2,27	1,05	11,09	19,45	0,43	3,09
P51: Nose #8	3,37	1,33	10,47	18,84	0,32	3,34
P52: Nose #9	2,95	1,5	8,94	18,04	0,35	1,52
P53: Nose #10	3,66	1,48	10,91	19,9	0,26	1,51
P54: Nose #11	3,07	1,31	12,08	18,95	0,58	4,25
Mean	**3,42**	**1,50**	**10,91**	**18,87**	**0,43**	**3,34**
Range	2,02	2,15	4,24	5,66	0,83	8,02

The average intensity of the expression “Happy” being detected was found to be 3.42, with the HAI being 4.29, presented in the “Nose #1” participant, and the LAI being 2.27, presented in the “Nose #7” participant. A range of 2.02 was found in the “Happy” expression intensity when altering the nose.

The average intensity of the expression “Sad” being detected was found to be 1.50, with the HAI being 2.75, presented in the “Nose #3” participant, and the LAI being 0.6, presented in the “Nose #0.6” participant. A range of 2.15 was found in the “Sad” expression intensity when altering the nose.

The average intensity of the expression “Angry” being detected was found to be 10.91, with the HAI being 13.19, presented in the “Default” participant, and the LAI being 8.94, presented in the “Nose #9” participant. A range of 4.24 was found in the “Angry” expression when altering the nose.

The average intensity of the expression “Surprised” being detected was found to be 18.87, with the HAI being 21.85, presented in the “Default” participant, and the LAI being 16.19, presented in the “Nose #5” participant. A range of 5.66 was found in the “Surprised” expression intensity when altering the nose.

The average intensity of the expression “Scared” being detected was found to be 0.43, with the HAI being 0.99, presented in the “Nose #5” participant, and the LAI being 0.16, presented in the “Default” participant. A range of 0.83 was found in the “Scared” expression intensity when altering the nose.

The average intensity of the expression “Disgusted” being detected was found to be 3.34, with the HAI being 8.97, presented in the “Nose #5” participant, and the LAI being 0.95, presented in the “Default” participant. A range of 8.02 was found in the “Disgusted” expression intensity when altering the nose.

5.2.7.1 Table 5: Nose – Analysis

The focus on the ‘Nose’ facial feature is noted to be less a focus on the nose itself, and rather a focus on the nasal area as a whole. This suggests that when interpreting the effect that the various nose shapes have on the emotional expression intensities, the focus should fall on how the area around the nose, such as the spaces between the eyes and area between the nose and mouth, are affected.

The nose shapes that are thinner and longer in shape (P45, P46) were noted to have Higher Average Intensities of ‘Happy’, while the fatter and shorter nose shapes (P50, P52) had Lower

Average Intensities of ‘Happy’. This difference is due to ‘Happy’ being interpreted from how wrinkled and contracted the lower nasal area (closer to the mouth) becomes. The longer noses would be interpreted by face reader as the mouth raising higher, even if the nose is what is changed, causing an increased intensity of said emotional expression. When looking for ‘Sad’, FaceReader would focus more on how the nasal area contracts around the eyes and close to the bridge of the nose. Nose shapes with thinner bridges (P49, P50) would produce a HAI of ‘Sad’ due to the contraction of the eyes being more visible, while nose shapes with wider bridges (P45, P46) would lean towards a LAI of ‘Sad’ as the contraction of the eyes would be less visible.

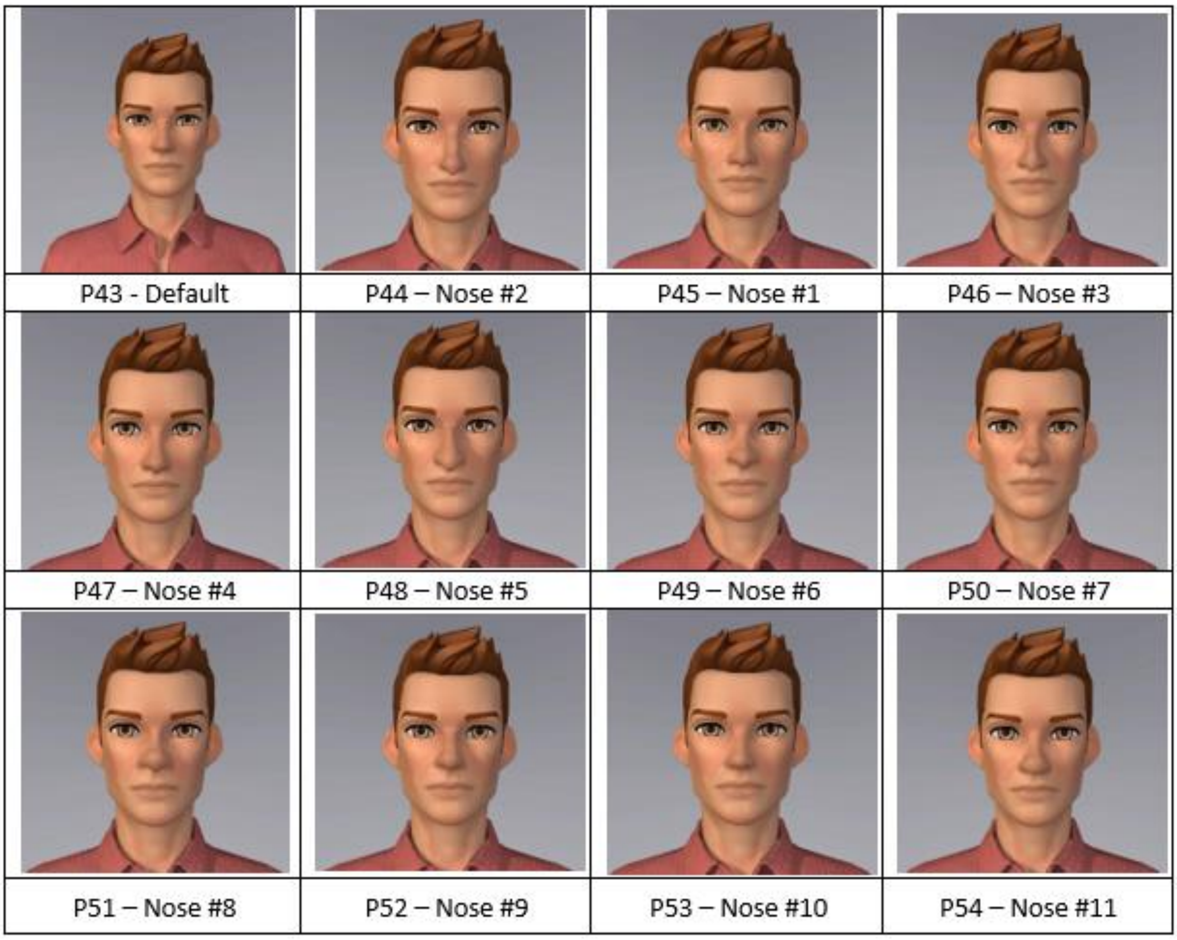


Figure 11 The ‘Ear’ participants.

A more intense ‘Anger’ expression would be found by FaceReader when there are more wrinkles around the nasal area due to a contraction of the muscles of the face. This is seen in the nose shapes that have wider lower areas (P43, P54) producing a HAI as take up less space around the eyes, which shows their contraction, and more space closer to the mouth, which creates the illusion of the mouth contracting more. Additionally, the nose shapes that are longer and more rectangular (P44, P52) lack these features and produce a LAI. “Surprised’ had an interesting result, as the intensity of this expression is measured in the nasal area relaxing and moving outwards, rather than contracting. The LAI of this expression was found in nose shapes that had a long and beak-like nose (P48, P49) while the HAIs were found in nose shapes that were rounded and less sharp. The corelation between the nose shape and the 'Surprise' expression is found to be whether or not the nose is rounder (almost smooth) in shape.

Inverse to the results of 'Surprise', the beak shaped noses were found to produce a HAI of the 'Scared' emotional expression. This is due to 'Scared' being interpreted by FaceReader through the wrinkling of the lower nasal area and a contraction of these muscles. The beak-shaped noses (P48, P49) put more emphasis on the lower nasal area, creating an illusion of contraction, while the LAI of 'Scared' was found in the rounder and shorter nose shapes (P43, P45). The emotional expression 'Disgust' had very similar readings to 'Scared', as the beak-shaped noses (P48, P49) had extremely high Average Intensity readings, when compared with the other 'Disgust' readings, and the rounder-shorter nose shapes (P43, P45) had much lower 'Disgust' Average Intensity readings. The reasoning for this is that the illusion of contraction on the lower nasal area and the more visible eye contractions allow for the face contractions associated with 'Disgust' to be identified by *FaceReader*.

5.2.7.2 Table 5 - Summary

The nasal area is one of the most important areas that is used by FaceReader when distinguishing the intensities of emotions. Rather than the actual nose being the focal point of analysis, how the nose's shape affects the nasal area was found to be the most important factor when choosing the nose shape. The length and width of the nose would produce different 'Happy' intensities, with a longer and thinner nose leaning towards a HAI, while the side of the nose bridge would affect the average intensity of 'Sad' being expressed. 'Angry' is more easily identifiable when a triangle-shaped nose is used (larger base, smaller bridge) as opposed to a rectangular nose (even size bottom and top) as the larger area around the eyes would assist *FaceReader* when identifying the contraction of the eyes. Additionally, the shape of the nose was found to be important in how a more pointed noise (similar to a bird beak) would produce a lower Average Intensity of 'Surprise', while a rounded and upwards pointed nose would produce a lower Average Intensity of 'Surprise'. The inverse of this is true for both the 'Scared' and 'Disgust' emotional expressions, as these expressions would increase with the illusion of contraction of the lower nasal area that the beak-like nose provides.

5.2.8 Mouth Testing Group

The average intensity of the expression "Happy" being detected was found to be 3.75, with the HAI being 4.55, presented in the "Lower Upper Narrow" participant, and the LAI being 3.25, presented in the "Upper Narrow" participant. A range of 1.3 was found in the "Happy" expression intensity when altering the nose.

The average intensity of the expression "Sad" being detected was found to be 0.87, with the HAI being 1.32, presented in the "Default" participant, and the LAI being 0.42, presented in the "Lower Upper Narrow" participant. A range of 0.9 was found in the "Sad" expression intensity when altering the nose.

Table 6 Mouth

Test 6: Mouth	Happy	Sad	Angry	Surprised	Scared	Disgusted
P55: Default	3,79	1,32	13,19	21,85	0,16	0,95
P56: Mouth #1	3,26	0,83	7,53	22,42	0,37	16,24
P57: Mouth #4	3,56	0,97	8,72	20,47	0,17	4,01
P58: Upper Narrow	3,25	0,62	11,21	19,75	0,31	0,66
P59: Lower Narrow	4,38	0,6	12,49	19,01	0,21	1,74
P60: Upper Wide	3,32	1,17	8,85	19,96	0,22	0,32
P61: Lower Wide	3,56	0,94	10,08	21,41	0,22	5,19
P62: Lower Upper Nar	4,55	0,42	11,22	16,97	0,16	0,49
P63: Lower Upper Wide	3,78	0,96	7,42	21,93	0,28	18,61
P64: Upper Narrow	3,37	1,27	8,96	20,38	0,14	0,49
P65: Lower Wide	4,1	0,69	10,11	21,23	0,13	2,94
P66: Wide Ext	4,04	0,61	9,8	22,21	0,29	5,3
Mean	**3,75**	**0,87**	**9,97**	**20,63**	**0,22**	**4,75**
Range	1,3	0,9	5,77	5,45	0,24	18,29

The average intensity of the expression "Angry" being detected was found to be 9.97, with the HAI being 13.19, presented in the "Default" participant, and the LAI being 7.42, presented in the "Lower Upper Wide" participant. A range of 5.77 was found in the "Angry" expression when altering the nose.

The average intensity of the expression "Surprised" being detected was found to be 20.63, with the HAI being 22.42, presented in the "Mouth #1" participant, and the LAI being 16.97, presented in the "Lower Upper Narrow participant. A range of 5.45 was found in the "Surprised" expression intensity when altering the nose.

The average intensity of the expression "Scared" being detected was found to be 0.22, with the HAI being 0.37, presented in the "Mouth #1" participant, and the LAI being 0.13, presented in the “Lower Wide” participant. A range of 0.24 was found in the “Scared” expression intensity when altering the nose.

The average intensity of the expression “Disgusted” being detected was found to be 4.75, with the HAI being 18.61, presented in the “Lower Upper Wide” participant, and the LAI being 0.32, presented in the “Upper Wide” participant. A range of 18.29 was found in the “Disgusted” expression intensity when altering the nose.

5.2.8.1 Table 6: Mouth – Analysis

The mouth, along with the eyebrows, is one of the most important areas of the face that *FaceReader* uses when determining emotional expression intensities. The positioning of the outer areas of the mouth and how wide the middle areas are spread apart are the main factors that are used for expression analysis, with the other areas of focus being the effect the mouth has on the cheeks and the lower nasal area.

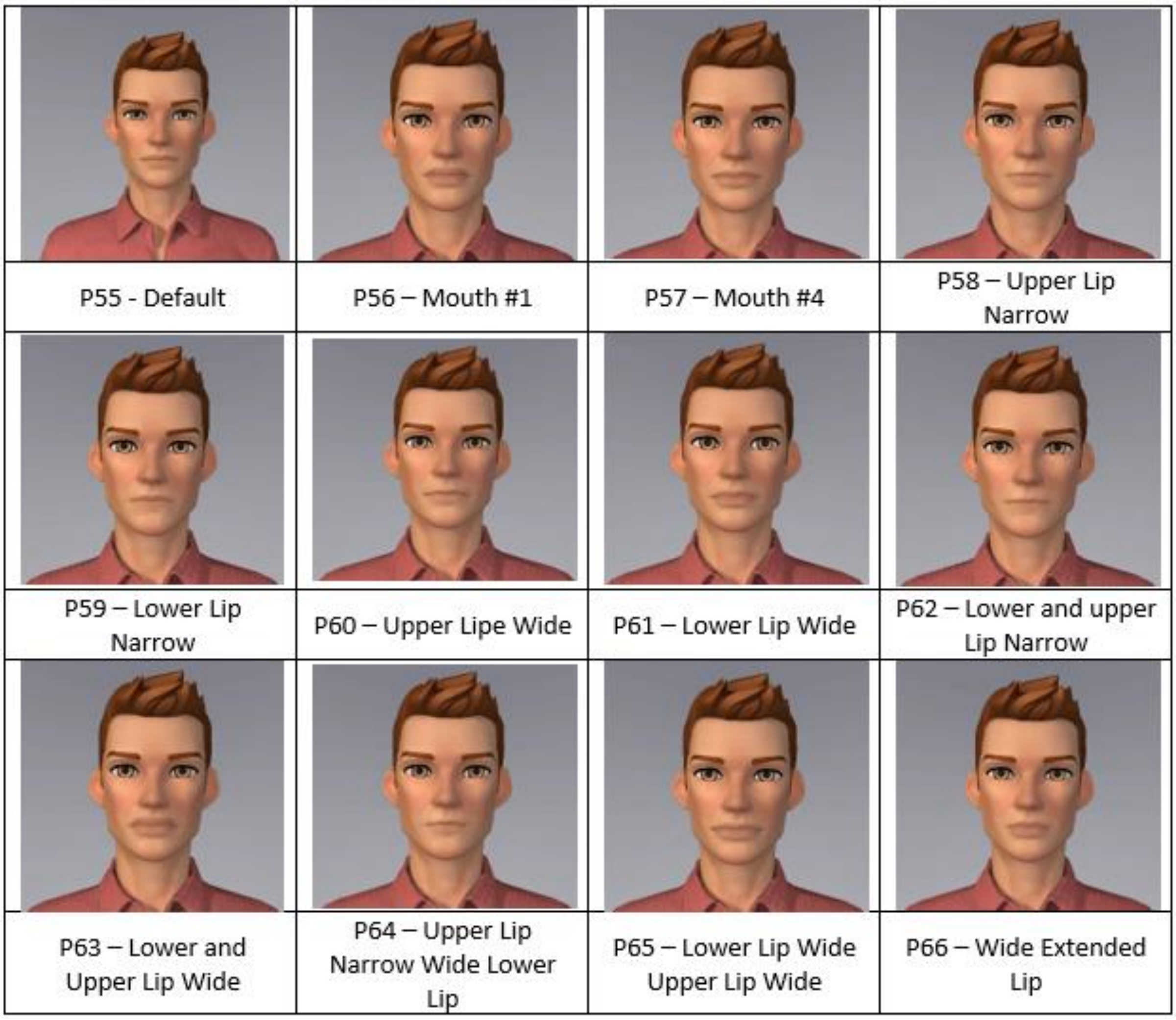


Figure 12 The ‘Ear’ participants

The intensity of ‘Happy’ is measured through how the corners of the lips are drawn up and higher on the cheeks, with a more intense expression being read when the teeth are exposed. When measuring ‘Happy’, the thinner lips (P59, P62) were found to produce the HAIs, while the thicker lips (P56, P58) were found to produce the LAIs. The noted differences in average intensities here would be due to the thinner lips having the ability to show more of the virtual avatar’s teeth when compared with thicker lip mouth shapes. The expression of ‘Sad’ is measured through how low the corners of the lips have dropped, which was noted to be more visible to FaceReader when the mouth shapes begin in a more neutral, or even slightly raised corner, position (P55, P60). This was found to be more apparent when comparing the HAIs to the LAIs, which found that the mouth shapes which would begin with the corners already slightly lowered (P62, P66) would be viewed as less ‘Sad’. This was likely due to the continuous calibration of FaceReader removing that facial feature’s personal bias towards being ‘Sad’.

When interpreting ‘Anger’ we would view how the corners of the mouth have dropped and how tense the lips are, as the lips could be tightly pressed together (physical anger) or open wide in a square shape (verbal anger). ‘Anger’ was viewed to have higher average intensities when using thinner and longer mouth shapes (P55, P59) while lower average intensities were viewed in thicker and compressed mouth shapes (P56, P63). These readings were likely due to FaceReader not being able to interpret the tensed movements of the thicker mouth shapes, whereas the thinner mouth shapes would easily show the drop of the corners and the tense movements in the lips. The emotional expression of ‘Surprise’ is the only noted emotional expression where the mouth is not seen as tensed, as the mouth would open wide and reveal the teeth in a state of shock. ‘Surprise’ was found to have a higher average intensity in the larger mouth shapes, in both thickness and length (P56, P66), while the lower average intensities were found in mouth shapes that are thinner and smaller (P59, P62). The difference in readings here would be due to the larger mouths being read as wide open, while the smaller mouths would be viewed as less open in comparison, leading FaceReader to view the larger mouths as being more ‘Surprised’.

While ‘Scared would actually look for the mouth to be more tensed and drawn back when compared with a ‘Surprised’ mouth shape, the intensity reading patterns for the mouth shapes were similar in these expressions. Mouths that were wider and thicker (P56, P66) were interpreted as being more ‘Scared’ while the lower average intensities for this emotional expression were again found in mouth shapes that are thinner and smaller (P59, P62). For similar reasons, FaceReader would interpret ‘Scared’ through how wide open the mouth was and how the lips have been tensed, meaning that the thicker lip shapes would lean towards being more ‘Scared’. Additionally, there are two mouth shapes that produced extremely high readings of ‘Disgust’; these mouth shapes were very similar as they both were compressed and thick mouth shapes (P56, P63). The lower average intensities were found in compact but very thin lip shapes (P60, P64) which is interesting as the intensity of ‘Disgust’ was found to only corelate to the thickness of the lips, rather than the length. This would make sense as FaceReader would look for ‘Disgust’ intensities through interpreting how the lower lip is tensed into the upper lip and how much the lower lip has protruded outwards.

5.2.8.2 Table 6 - Summary

The mouth is an extremely important part of Facial Expression Recognition as every emotional expression has an effect on the movement of the mouth. The thickness and length of the mouth were found to be the most important factors that FaceReader looks for when interpreting the expression intensities. Thinner lips would produce a higher than average ‘Happy’ reading, while thicker lips would be corelated with ‘Disgust’. The length of the mouth (measured from corner to corner) would change how FaceReader interprets the movement of the corners, as a longer mouth produced higher ‘Angry’ intensities, while a shorter and thicker mouth was associated with both the ‘Surprised’ and ‘Scared’ emotional expressions. Additionally, the ‘Sad’ emotional expression readings indicated that the starting position of the mouth was also an important factor, as a mouth with a default position that has lower corners of the mouth would skew FaceReader’s interpretation of ‘Sad’.

5.2.9 Analysis: Ranges

Table 7 Ranges

Feature	Happy	Sad	Angry	Surprised	Scared	Disgusted	Sum of the ranges
Face Shape	1,83	0,82	5,35	6,98	0,32	1,18	16,48
Eyebrows	1,43	1,14	9,59	7,83	0,2	1,38	21,57
Eyes	1,34	5,84	7,85	11,96	0,41	1,39	38,05
Ears	2,46	1,35	5,88	3,45	0,15	1,39	14,68
Nose	2,02	2,15	4,24	5,66	0,83	8,02	22,92
Mouth	1,3	0,9	5,77	5,45	0,24	18,29	37,6

Table 7 contains the values from the ‘Range’ row found in tables 1-6. As stated previously, the range was found by subtracting the Lowest Average Intensity value from each emotion expression column from the Highest Average Intensity value in that same column. The resulting value will indicate the amount of difference that each facial feature would have on FaceReader’s emotional expression analysis. This would be noted in that a high range value would only be possible if there is a noticeable difference between the Highest Average Intensity value and the

Lowest Average intensity value. This analysis will be done by going through each column, to identify the facial feature that had the largest impact on that emotional expression, and the final column will be discussed through looking at which facial feature had the highest sum of ranges. By investigating the sum of the ranges for each facial feature, we would be able to prove the importance of those facial features in expression analysis as a higher sum of ranges would indicate a higher control over emotional intensities.

5.2.9.1 Table 7 –Summary

The highest ranges of intensity averages for the 'Happy' emotional expression were found within the 'Ears' and 'Nose' facial features, while the lowest ranges were found in the 'Mouth' and 'Eyes' facial feature group. The importance in these results is that the ears and nose are not focal areas when measuring the intensity of 'Happy', while the mouth and eyes are the most important areas of focus. This indicates that *FaceReader's* expression recognition would have consistent intensity readings when the assumed to be more important facial features vary in shape and size. Additionally, the higher ranges caused by the changes of shape and size of the ears and nose would likely be attributed to how they affect the visibility of the other facial features. For example, larger ears would alter how well FaceReader can identify the crows-feet appearing in the corner of the eyes and a longer nose would alter the visibility of the wrinkling of the nasal area which is used to identify the intensity of 'Happy'.

When analysing the 'Sad' emotional expression, there is a noticeably higher average intensity range in the 'Eyes' testing group while the lower average intensity ranges were produced by the 'Face Shape' and 'Mouth' testing groups. The reasoning for this is due to an outlier in the eyes, where one eye shape was shown to produce a much higher 'Sad' reading, but this information is still valuable as it indicates that the shape of the eye will cause a change in how the virtual avatar's emotional expressions are identified. This corelates with our understanding of how 'Sad' is expressed through facial features as there is a large focus on the angling of the eyes when determining the intensity of the sadness being portrayed. While there are fairly lower ranges, found in the face shapes and mouths, the other facial feature groups produced very similar 'Sad' average intensity ranges. This is likely due to the variations within these facial features would not change the intensity of sadness being portrayed as there is a lower focus on intensity from these facial feature areas.

The 'Eyebrows' testing group produced a notably higher average intensity range when portraying 'Angry', which would also fit into our understanding of how this emotional expression is portrayed. The length and angle of the eyebrow can greatly change the appearance of ‘Angry’ on the face, as this is the facial feature that *FaceReader* was found to focus on most when identifying the intensity of the subject’s anger. The ‘Nose’ testing group produced the lowers average intensity range as the nasal area is not a focal point for *FaceReader* when determining ‘Angry’. What was interesting in the ‘Angry’ readings is that the ‘Face Shape’, ‘Ears’, and ‘Mouth’ testing groups all had very similar ranges. This corelates with how this emotional expression is interpreted and analysed as these facial features were found to not be important in *FaceReader’s* analysis of the intensity of ‘Angry’.

The emotional expression of ‘Surprise’ was found to have a notably higher average intensity range in the ‘Eyes’ testing group, with the ‘Eyebrows’ testing group the second highest fairly lower. The lower average intensity ranges were found in the ‘Ears’ testing group, with the ‘Nose’ and ‘Mouth’ testing groups being only slightly higher. These ranges would also corelate with our understanding of how FaceReader interprets emotional expressions, as the intensity of ‘Surprise’ is interpreted through how much of the iris is visible. This would mean that when the eye shape changes, the visibility of the iris changes drastically as well. The other facial features, such as the ears and nose, are less important as they are not noted as important for the analysis of the intensity of ‘Surprise’. What was interesting to note was that the mouth was found to not have that much of an effect on the intensity ranges of the ‘Surprise’ emotional expression, which contradicts the understanding that a wider opened mouth would be viewed as more intense. What this tells us is that FaceReader does not elevate the intensity of ‘Surprise’ if there are smaller variations in how wide the mouth has been opened.

The ‘Nose’ testing group produced a much higher average intensity range (more than double the next highest range) when the intensity of ‘Scared’ was analysed by FaceReader. This is understandable as the different nose shapes and sizes would affect how contracted the nasal area would become as a more intense feeling of fear would contract around the nasal area, while a less intense fear would be more relaxed. This would indicate that when measuring how intense the avatar’s ‘Scared’ expression is, *FaceReader* puts a much higher focus on the nasal area than what was initially thought. The ‘Ears’, ‘Eyebrows’, and ‘Mouth’ testing groups produced the lower average intensity ranges due to *FaceReader* finding these facial features to have a lower impact on the intensity of ‘Scared’. The ‘Eyes’ testing group had the second highest average intensity range as *FaceReader* would also look for the visibility of the iris when determining the emotional expression being displayed, although these results determined that FaceReader does not focus as much on the iris visibility when analysing ‘Scared’ when compared with when FaceReader analyses ‘Surprise’.

The average intensity range of ‘Disgust’ was found to be much higher in the ‘Mouth’ testing group when compared with the other testing groups. The ‘Nose; testing group was the next highest range, yet still less than half as large, and the other four testing groups produced remarkably similar ranges. The higher average intensity range of ‘Disgust’ produced through the different mouth shapes indicates that *FaceReader* places a high focus on the positioning and contraction of the mouth when analysing this emotional expression. ‘Disgust’ is known to be found through how the face would contract towards the centre of the face, and this contraction is mainly found through analysing the mouth and nasal areas. This also provided evidence for why the ‘Nose’ is the only other facial expression that is a focal point for *FaceReader’s* ‘Disgust’ intensity readings, whereas the other four facial feature groups will not have as much of an effect. The extremely similar average intensity ranges of the ‘Eyebrows’, ‘Eyes’, and ‘Ears’ is what is most interesting as the eyes and eyebrow areas were found to have a much lower impact on the intensity of ‘Disgust’ than what would be assumed.

5.2.9.2 Table 7 –Conclusion

Through comparing the average intensity ranges of FaceReader’s expression recognition component, it was found that *FaceReader*, and likely other computer vision programs that use similar frameworks, will focus heavily on one area of the face when measuring the intensity of an emotional expression. To clarify, the presence of an emotional expression was consistent regardless of changes to the facial features as there was always a noted presence of each of the measured emotional expression. The intensity of these emotional expressions, however, would have noteworthy variations depending on which facial feature was changed and what changes occurred.

The ‘Ears’ testing group had the highest effect on the intensity reading of the ‘Happy’ emotional expression. Theoretically, the ears should not affect how ‘Happy’ the virtual avatar would look as the ears are not noted as important for facial expressions. However, the variation in intensities would be due to the *Animaze* choices of ears allowed the changing of both the ear shape and size, from extremely small to extremely large, which would overlap with how FaceReader would analyse the outer corners of the eyes (crows-feet area) and the width of the cheeks. These areas will affect the intensity of ‘Happy’ as they are noted as more important for differentiation by FaceReader. To avoid the potential influence that the ears will have on facial expression recognition, animators would need to create a clear distinction between where the outline of the face ends and the ears begin. This is best done through the lower area of the ears (the earlobe) being prominent and the upper ear area facing out and away from the side of the face.

The ranges found within the 'Sad' emotional expression intensities were closely linked to the 'Eyes' testing group. The shape and size of the eye was found to be much more important than any other facial feature when FaceReader interprets the intensity of the emotional expressions. The reasoning for this is that the intensity of 'Sad' was found to mainly be measured through the angling of the eyes, and how visible the eyelids are. Eye shapes that are predisposed to being angled with the outer corners of the eyes going down and the inner corners moving up would produce lower sadness intensities as *FaceReader* would calibrate these eye shapes to a neutral position which leans toward 'Sad' to remove personal bias. Additionally, *FaceReader* would interpret the positioning of the eye lids to determine intensities, with the visibility of the iris decreasing with more intense 'Sad' expressions. With these factors in mind, when designing a virtual avatar to lean towards looking 'Sad', it is necessary to have larger eyes that display the eyelid control fully, and the eyes would need to begin in either a straight line from one corner of the eye to the other or alternatively they can begin angled inwards with the outer corners raised.

When measuring the average intensity ranges of 'Angry', the highest variation in intensities was found in the 'Eyebrows' testing group. Animaze had a large variety of eyebrows to use with the variations being in the thickness, angles, and how rounded they were. *FaceReader* would measure the intensity of the 'Angry' expression through the curvature of the eyebrows and their thickness as the intensity of anger in facial expressions is noted through how drawn together the eyebrows are and how low the inner corners of the eyebrows drop. This meant that the eyebrow shapes that were flat and thinner would produce higher intensities when compared with the arch-shaped and thicker eyebrows. To lean toward an 'Angry' expression with a virtual avatar, the recommended eyebrows to use would be flat from one end of the eyebrow to the other (no rounding) and they should be fairly thin in size (not bushy) to emphasize the downwards movement of the inner eyebrow area.

The 'Eyes' testing group was found to be important in the intensity ranges of another emotional expression as this facial feature group produced a notably higher variation of ranges when measuring 'Surprised'. For similar reasons to how the intensity of the 'Sad' emotional expression is measured, the importance that *FaceReader* places on the visibility of the iris leads to the intensity of 'Surprised'. The *Animaze* eye shapes which were larger in size produced much higher 'Surprised' intensities when compared with the smaller eye sizes. The eye size is the most important factor as the larger eyes were able to show more of the iris, which *FaceReader* interpreted as a more intense expression of 'Surprised'. The angle of the eye was found to be less important for this emotional expression when compared with the 'Sad' emotional expression intensities as the intensity readings of ‘Surprised’ were consistent regardless of the angling of the eyes, yet slightly higher readings were found in flat angled eye shapes. These findings indicate that when creating a virtual avatar that would have *FaceReader* lean toward analysing higher intensities of the ‘Surprised’ emotional expression, it would be necessary to use larger eye sizes (that can display the iris easily) that are angled evenly from corner-to-corner.

An analysis of the ‘Nose’ testing group found a much higher range of emotion expression average intensities when *FaceReader* analyses the intensity of ‘Scared’. The affect that the nasal area has on the intensity of ‘Scared’ was attributed to how the shape and size of the nose alters *FaceReader’s* view of the nasal areas around the eyes and around the mouth. The size and thickness of the bridge of the nose (the area between the inner eye corners) would alter the visibility of the contraction of the eyes, which alters the intensity of ‘Scared’ as less tensed eye contractions can be seen as higher fear. Additionally, the larger the lower area of the nose (around the nostrils) will be interpreted by *FaceReader* as a contraction of the lower nasal area causing a higher intensity reading of ‘Scared’. With these factors in mind, when aiming to create a virtual avatar that leans towards a higher intensity of the ‘Scared’ emotional expression it would be best to use a beak-shaped triangle nose. This nose would have a thin bridge and a larger lower area, which creates a triangle shape, with a pointed and curved tip of the nose which covers more of the lower nasal area.

When analysing the emotional expression ‘Disgust’, there was an extremely high intensity average range in the ‘Mouth’ testing area when compared with the other facial feature groups. The mouth was found to be the focal point of FaceReader when determining the intensity of ‘Disgust’ as a face that is expressing more of this emotion would have a protruding lower lip and a wrinkled nasal area between the nose and upper lip. This was confirmed in the thicker lip shapes with shorter widths (where thickness is from the top of the lip to the bottom of the lip and the width is from corner-to-corner) producing much higher ‘Disgust’ intensity averages. The thinner mouth shapes, especially those with thinner upper lips, that were wider produced much lower ‘Disgust’ intensity averages. With these factors in mind, to design a virtual avatar that leans towards higher intensities of ‘Disgust’ it is extremely important to use a mouth shape that has thicker lips and are shorter in width. This mouth shape would have *FaceReader* lean towards higher ‘Disgust’ intensities, as there would be a perceived further protruding of the lips that would also wrinkle the nasal area between the nose and upper lip.

5.2.10 Facial Feature Focus – Conclusion

To further understand which facial features had a large impact on emotional expression intensities, the following section shifts the analysis focus from the facial features to the emotional expressions.

5.3 Results: Expression Focus

The following section will be analysing the same results presented in Tables 1-6 but with the data now grouped to focus on the emotional expressions rather than the facial features. This will be done by grouping the emotional expression results above and presenting them as a bar graph. By placing the average intensity results all on the same graph, it will be easier to visually distinguish differences in the results through visual comparison and an overall average can be found. Bar graphs were chosen as they would provide the clearest visual distinctions between the intensity readings as different sized data bars will represent higher or lower average intensities.

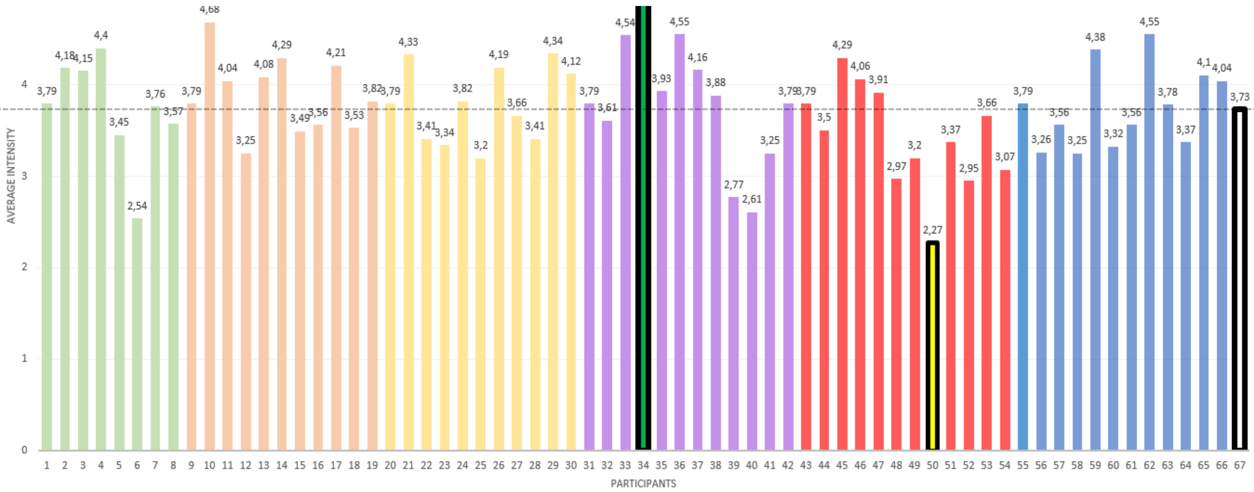
The following bar graphs can be read as follows:

- The Y axis is Average Intensity reading taken from the above tables.
- The X axis is the participant number from the above tables. This will clarify which facial feature the average intensity belongs to.
- P1-8 (green) is the 'Face Shape' group, P9-19 (orange) is the 'Eyebrows' group, P20-30 (yellow) is the 'Eyes' group, P31-42 (purple) is the 'Ears' group, P43-54 (red) is the 'Nose' group, and P55-66 (blue) is the 'Mouth' group.
- P67 (white with black outline) is the average of the P1-66 readings, and this average is further represented by the dotted line that runs across the graph.
- The green bar with a black outline is the Highest Average Intensity across all the participant results.
- The yellow bar with a black outline is the Lowest Average Intensity across all the participant results.

As these graphs visualise the results from the facial feature groups, the intent is to use them to further investigate which facial feature areas are most important for FaceReader when identifying the intensities of facial emotion expression. This will be done by first summarising the graphs from a glance value perspective, where the groups will be compared with the overall average (the dotted line), which will be followed by an analysis of the Highest Average Intensity and Lowest Average Intensity. This analysis will be done through a comparison of the rest of the data in an attempt to identify outliers and perhaps the facial feature group that has the highest effect on the emotional expression.

5.4.1 Happy Expression Focus

Table 8 Happy Expression Graph Results



5.4.1.1 Happy Graph – Analysis

The first graph presented here represents all of the average intensities of the 'Happy' emotional expression. As mentioned previously, the intensity of happiness is examined from a human perspective through mainly the lower area of the face. There are few distinct brow movements, and the focus on the eyes is on contractions causing 'crows-feet' that is caused by a raising of the cheeks, but there are important areas to note around the mouth area. The raising of the corners of the lips will be identified as a higher intensity of happiness, which will be further emphasised by a visibility of the teeth as the mouth opens with a large smile. When looking at the intensity bar graph one would notice that across all the participants, the intensity of 'Happy' is fairly consistent. This can be said as there were few noteworthy outliers in both the higher and lower intensity readings. The higher readings stayed close to 4.5, with the Highest Average Intensity being 5.07 (P34), and there were several readings close to the Lowest Average intensity of 2.27 (P50).

P10 – Eyebrow #1	P34 – Front Facing Medium	P6 – Square	P50 – Nose #7

Figure 14 Notable 'Sad' facial features

Each facial feature group shared similarities in the peak of their intensities, although the 'Eyebrows' and 'Ears' facial feature groups presented the most consistently high readings with the presented averages of 3.89 and 3.83 respectively. On the other end of the 'Happy' readings, the lowest average was found in the 'Nose' facial feature group which averaged at 3.42. These results would indicate that these facial features are the most important when designing a virtual avatar that will lean-toward or shy-away from a 'Happy' expression. The 'Ears' were said to not be of importance for any facial emotion expression, yet previous discussions attributed their impact to the way the shape and size of the ears may alter how FaceReader interprets the outline of the face. When this happens, the intensity of 'Happy' would drop as the larger surface area around the cheeks will make *FaceReader* assume the cheeks have not risen high enough for a more intense expression. Additionally, the 'Eyebrows' would be consistently higher as there are six eyebrow shape options in *Animaze* that are slightly rounded (arc-shaped). This shape of eyebrow would cause a bias towards 'Happy' due to FaceReader recognising this shape as there being no contraction around the eyes, a movement commonly associated with other negative emotions, creating a false "Happy'.

Conversely, the 'Nose' facial feature group's impact on the intensity of 'Happy' was discussed above and it was found that a nose that is longer and rectangular in shape (similar sized bridge and lower area) produced a higher 'Happy' intensity as the longer noses created an illusion of a higher raised upper lip that *FaceReader* interprets as a higher intensity. It was found that Animaze only had three nose shapes that matched this criterion, with the rest either having a shorter length or were triangular in shape (thin bridge and large bottom). What was also found to be interesting is that the averages for the other facial feature groups were very close to the overall average (3.73) of 'Happy'. 'Face Shape' had the average of 3.73, 'Eyes' averaged at 3.78, and the 'Mouth' group had the average of 3.75. This information tells us that *FaceReader's* ability to recognise and analyse the intensity of 'Happy' will be consistent even when these facial features are changed. There may be slight changes in the intensities when there are huge changes, such as extremely small eyes, thick lips that lack corners, or a jaw shape that creates too much surface area around the mouth, but these facial feature groups were found to not be of great importance to *FaceReader* when analysing ‘Happy’.

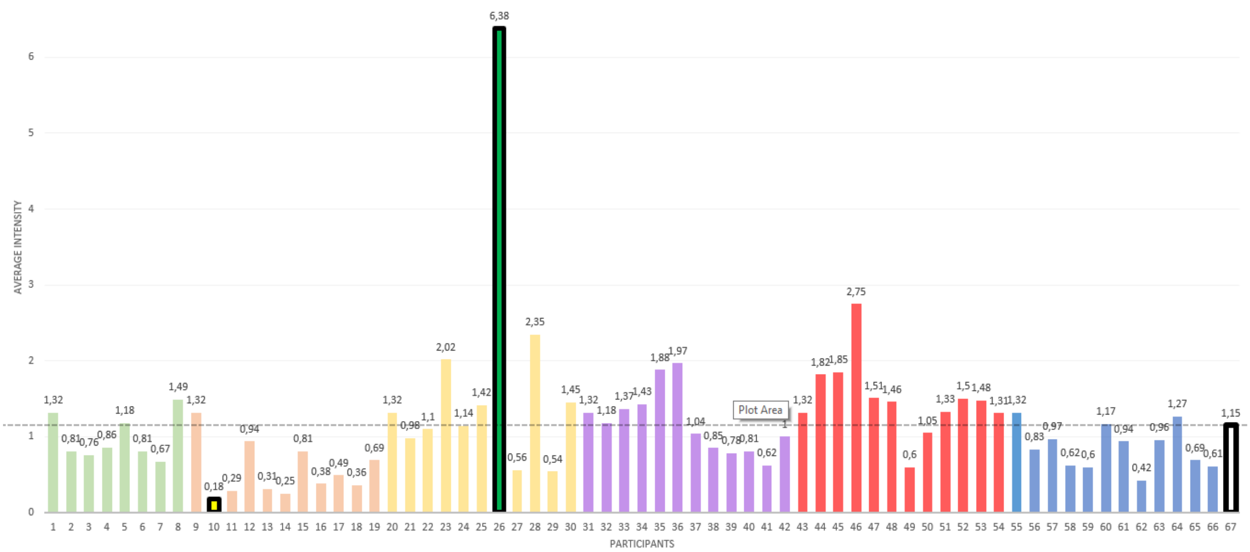
5.4.1.2 Happy Graph - Discussion

The average ‘Happy’ expression intensity of each facial feature group was found to be useful, but rather than being useful in finding which facial features have the highest impact on *FaceReader’s* expression recognition feature the value in these readings is in recognizing the limitations of *Animaze’s* editing choices. Animaze provides a surplus of facial features to choose from when editing virtual avatars, but what the above analysis found was that there are several of the same shape of facial feature present in each facial feature group. What was found through this discovery is that the averages of the expressions would be representative less of the impact of that facial feature and more-so a representation of which shape or type of that facial feature is more present.

As an example, the ‘Nose’ facial feature group had the lowest overall average of the ‘Happy’ intensity. The lower readings found when editing the note facial feature was due to there being a much higher selection of nose shapes that lean away from expressing happiness when compared with the amount of nose shapes that lean toward happiness. The opposite of this situation was true for the ‘Eyebrows’ group presenting more options of options that would lean towards higher ‘Happy’ intensities. This is not to say these facial features have a lower impact on expression recognition, rather that the range of values is more useful in proving their importance.

5.4.2 Sad Expression Focus

Table 10 Sad Expression Graph Results



5.4.2.1 Sad Graph – Analysis

The intensity of ‘Sad’ was shown to have a relation to the entire face, as the position of the mouth corners and height of the eyebrows both have an effect on how sad one looks, but the eyes are the focal point of this emotional expression. Sadness is known to be a hard expression to mask as even if one attempts to hide the expression with a smile, ‘sad-eyes’ are easily identifiable by a dropping of the eye corners and less iris visibility due to lowered eyelids. At a glance, the graph presenting all of the ‘Sad’ intensity readings presents an extreme outlier in the Highest Average Intensity reading (P26) which is a part of the ‘Eyes’ facial feature group. This reading was produced by the eye shape that was the smallest in size and had slightly lowered eyelids, which match the criteria for eye shapes that affect the intensity of sadness. The Lowest Average Intensity (P10) is less apparent as there are several readings that are close to this value, yet all these values were in the ‘Eyebrows’ feature group.

P26 – Small Monolid	P46 – Nose #3	P10 – Eyebrow #1	P14 – Eyebrow #5

Figure 17 Notable ‘Sad’ facial features

The lower intensities produced through the changes in eyebrow shapes would corelate with what was found in the ‘Happy’ graph analysis, as the eyebrow shapes that are provided by Animaze are rounded and provide higher ‘Happy’ intensities. This is further supported by how the few eyebrow shapes that are straighter in shape (less arc-like) provided intensity readings that were closer to the overall average. The overall average of the ‘Sad’ intensities was calculated as 1.15. Visually, one can see that each facial feature group leans towards having results that are either noticeably higher or lower than this average value. The ‘Ears’ were the closest group this average, being calculated to be 1.19, where it was apparent that the importance of the ear is how the shape and size affect FaceReader’s ability to detect where the face starts, and the ear begins. There were three groups that were below this overall average: the ‘Face Shape’ group averaged at 0.99, the ‘Eyebrows’ averaged at 0.55, and the ‘Mouth’ group averaged at 0.87. Each of these groups had several readings fairly far below the overall average line due to these groups having higher quantities of facial feature options that shared the ability to avoid being identified as ‘Sad’. The reasoning for the ‘Eyebrows’ options being perceived as less sad were mentioned previously, but several ‘Face Shape’ options had wider chins and square jaws which *FaceReader* interpreted as being less sad due to the cheeks dropping less. Additionally, the ‘Mouth’ options tended to be seen as less ‘Sad’ due to the majority of the lip shapes being thinner on the upper lip. This shape of lip reveals more of the teeth when the mouth is opened and would also have corners that are seen as dropping less when compared with mouth shapes with thicker upper lips.

The two groups that were higher than the overall average for ‘Sad’ were the ‘Eyes’, which averaged at 1.75, and the ‘Nose’, which averaged at 1.50. According to *FaceReader*, the choices of eyes that would lean towards being sadder are those that were smaller and had an upper eyelid that is already dropped and covers part of the iris. This criterion is matched by several eye shapes and is the reason for the average intensities in the ‘eyes’ testing group being higher than the other features. While not as high, the intensities of the ‘Nose’ choices were found to also produce several high readings as only two of the nose feature choices were below the overall average line. Higher ‘Sad’ intensities were produced when a nose that had a prominent, yet thin, bridge of the nose was used, and majority of the nose shape choices provided by *Animaze* had this description. The size of the nose bridge is important as an aspect of sadness intensities lies in how the eyes are contracted, so a larger bridge would disrupt *FaceReader’s* ability to analyses how contracted the eyes are. The nasal area is not intended to of extreme importance in the analysis of this emotional expression, but the way the nose shape deforms the area around the nasal is what causes changes in these readings.

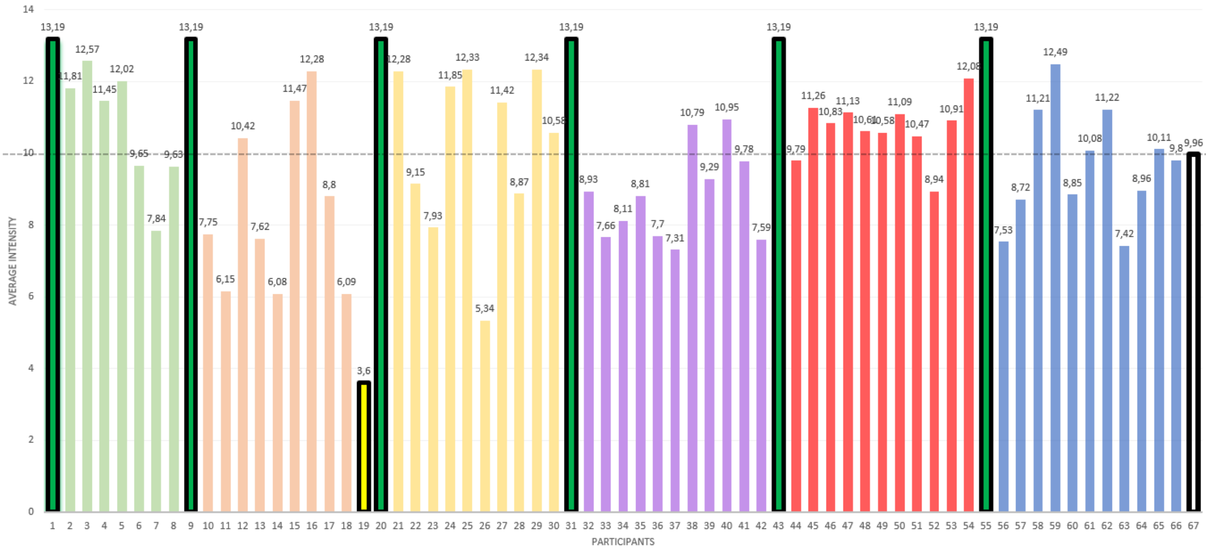
5.4.2.2 Sad Graph - Discussion

The impotence that FaceReader places on the ‘Eyes’ when it interprets the intensities of the ‘Sad’ emotional expression was further emphasised in this subsection. There were much lower readings overall in this graph as several of the facial feature groups contained the feature shapes that tended to avoid expressing sadness. This may also be less so due to an actual avoidance of sadness, but rather a further push into the problems *FaceReader* may face when interpreting these emotional expressions. The way that the ‘Nose’ and ‘Face Shape’ facial feature groups affect the intensity of sadness is less so due to their own shape and more to do with how they affect the visual clarity of the other facial features.

When FaceReader loses the ability to clearly see specific facial feature areas, it tends to lower the overall intensity of the emotional expression that is based off that area. In this example, the intensity of ‘Sad’ is affected by how FaceReader can see the contraction between the eyes. When a larger nose shape is used, FaceReader can no longer see a contraction and so it will lower the intensity of ‘Sad’ rather than compensating another way. With this in mind, the way facial features affect each other is just as important as the feature itself.

5.4.3 Angry Expression Focus

Table 120 Angry Expression Graph Results



5.4.3.1 Angry Graph – Analysis

‘Angry’ is expressed through every facial feature working together, but the intensity of the anger is found through how analysing close the eyebrows have been drawn down and together and the shape of the mouth’s opening. The above graph is immediately visually significant as it highlights how the default virtual avatar’s facial features produced the highest average intensity reading for the ‘Angry’ emotional expression. The reasoning for these six readings being the same is due to the first participant in each group (P1, P9, P20, P31, P43, P55) presenting the same facial features which produces the same reading as one another in each expression graph. The reasoning that the same avatar was used in each testing group was clarified in earlier sections, with this reasoning being that the default virtual avatar features are still noteworthy and will produce results that are significant, such as the ‘Angry’ intensities graph. Additionally, the lowest average intensity was detected as 3.6 (P19) and attributed to the unibrow shaped eyebrow. This reading was noticeably lower than the rest of the ‘Angry’ expression readings, indicating the effect the eyebrow shape has on the anger intensities.

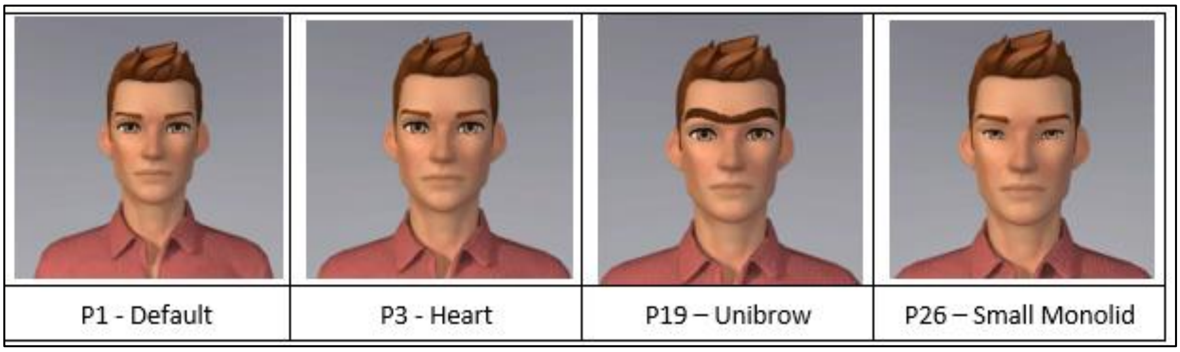


Figure 19 Notable 'Angry' facial features

The overall average for "Angry' intensities across all the participants was calculated as 9.6. The average of the 'Mouth' facial feature group was the closest to this, calculated as 9.97, with the 'Ears' having the next closest average at 9.18 and the 'Eyebrows' close by with an average of 8.50. The 'Eyes' and 'Nose' facial feature groups had similar, but higher, averages at 10.48 and 10.91 respectively. The facial feature group with the highest average was found to be the 'Face Shape' group with an average of 11.02. These averages and the presented graph indicate that FaceReader's focus on the eyebrows is prominent due to the range of intensities being quite high when the eyebrows are changed. The 'Angry' intensities fluctuated with the eyebrows, where when a rounder shaped eyebrow was used the intensity of anger would drop. This is likely due to there being less of a contraction and dropping of the eyebrows noticed by *FaceReader*. Additionally, when compared with the eyebrows, this information highlights the lack of importance that the other facial features have when *FaceReader* searches for the 'Angry' emotional expression.

The 'Face Shape' has a higher average intensity as the square-shaped and large jaw is the most prominent feature in the different 'Face Shape' choices, which were found to produce a higher intensity of anger. The 'Eyes' also fluctuated a lot in the readings they produced, with the smaller eyes producing lower anger intensities and the larger eyes producing higher intensities. This would be attributed to the visibility of the iris being a factor in anger intensity and the larger eyes would create a higher visibility for the lowering of the eyebrows. What was found in the 'Nose' facial feature group results was that the nose does not have a high contribution to the intensity of anger, as the readings were the most consistent in this testing group in comparison to the other testing groups, but there is a slightly lowered intensity reading when the nose has a significantly sharp and prominent bridge of the nose. In *Animaze*, there are more options of facial features that do not fit the above-mentioned 'Nose' shape criteria, which is why these groups had slightly higher averages when compared with the remaining facial features.

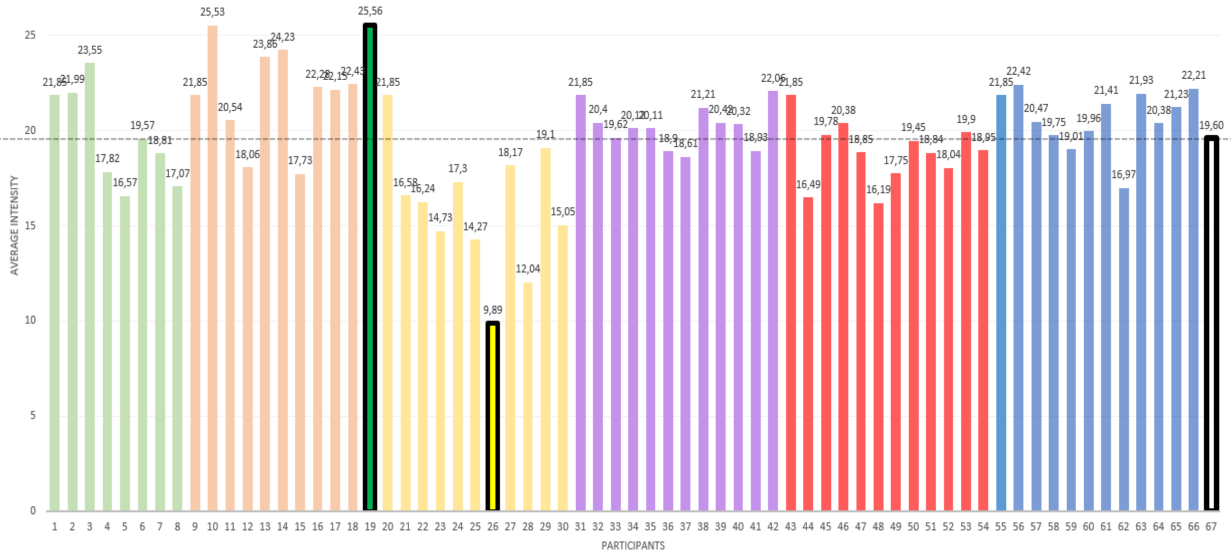
The ‘Mouth’ testing group’s overall average was slightly higher than the total average, and the ‘Ears’ testing group was slightly lower. Both of these facial features were extremely close to the overall average, yet the visualisation of their results provides interesting information. The ‘Ears’ are not intended to have an impact on the intensity of the ‘Angry’ expression, yet the ear shapes that protruded outwards and were easily identifiable provided higher intensities, whereas the smaller and flatter ears provided lower than average intensities. This effect is likely due to the ears altering how FaceReader interprets the ends of the eyebrow area and how they have risen. The ‘Mouth’ test results are seen as having an even number of results that are both higher and lower than the overall intensity average. This would indicate the importance that *FaceReader* places on the mouth when analysing the intensity of ‘Angry’ as the mouth shapes that had a thin upper lip and were shorter in width produced higher intensities when compared with the thicker lipped and wider mouth shapes. Both the shape and width had an effect on the intensity of the anger, but the importance was more so on the upper lip thickness, and as there were several mouth choices that used at least one of these criteria the results were highly varied.

5.4.3.2 Angry Graph - Discussion

The results from every ‘Angry’ intensity reading presented the fact that the default virtual avatar produced the highest average intensity across every facial feature choice. The default virtual avatar has no discernible facial features, with each feature being shaped to match an average American-Caucasian male in their early 30’s. The eyes are not too big or small, the ears fit neatly on the side of the head, the nose is not protruding or emphasised, the mouth’s lips are fairly thin but noticeable, and the eyebrows are average in thickness but shaped in a cartoon-like fashion. These details could suggest several ideas: *FaceReader* would create a bias in a neutral face to lean toward higher ‘Angry’ intensities, the default *Animaze* features are able to easily produce anger due to their range of motion, or the perfect virtual avatar for presenting ‘Angry’ expressions clearly is one that is evenly spaced out and provides a clear view of the rest of the face’s features. All of these statements are plausible, but the exact reasoning would require further testing which focuses on expressing anger through virtual avatars.

5.4.4 Surprised Expression Focus

Table 14 Surprised Expression Graph Results



5.4.4.1 Surprised Graph – Analysis

At a glance, the ‘Eyebrows’ facial feature testing group is the most effective at providing higher expression intensities of the ‘Surprised’ emotional expression. Surprise is portrayed through each facial feature, the widening of the eyes and relaxing of the nasal area, but the focus lies in how high the eyebrows have raised and how wide the mouth has opened due to the jaw dropping. As discussed through the ‘Angry’ graph analysis, there is a higher number of eyebrow shapes that are rounded in shaped (arched) which, when they rise in surprise, *FaceReader* would interpret as the virtual avatar expressing a higher intensity of ‘Surprised’. The higher number of this type of eyebrow is likely the cause for that testing group being noticeably higher when compared with the other facial features, and when compared with the overall average of 19.60.

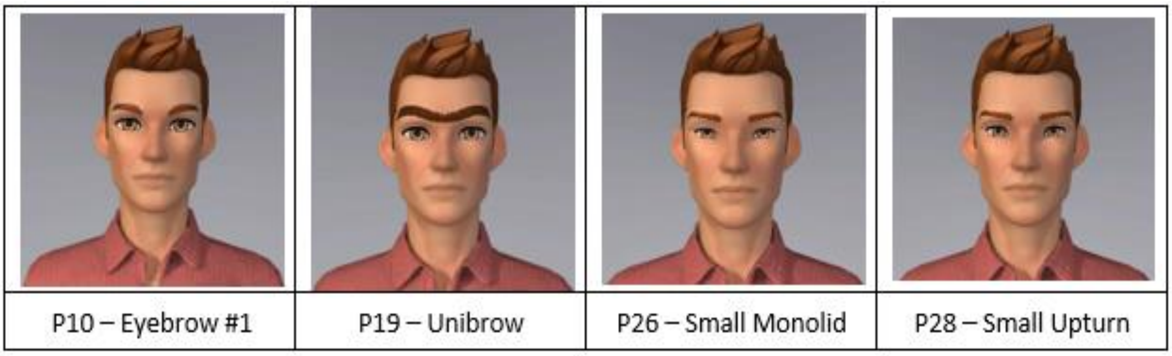


Figure 20 Notable ‘Surprised’ facial features

The highest average intensity of ‘Surprised’ was found as 25.56 (P19), produced by the eyebrow that is shaped as a unibrow with rounded sides, and the lowest was 9.89 (P26), the smallest eye shape. The highest intensity average was produced due to the presented eyebrow shape being extremely thick and rounded in shape, so when they rise *FaceReader* would interpret the virtual avatar as expressing extreme surprise. The lowest intensity average was produced as the small size of the presented eye shape prevented *FaceReader* from identifying a wide opening of the eyes, due to the lack of iris visibility. Both of these facial feature groups also presented the closest readings to the HAI and LAI, as the eyebrows produced the second highest average intensity (25.53, P10), and third and fourth highest, and the 'Eyes' produced the second lowest average intensity (9.89, P26) readings, and the third, fourth, and fifth lowest readings. As discussed previously, consistently high, or low, results indicate the importance that these facial features have when *FaceReader* identifies the intensity of that emotional expression.

The other facial feature testing groups were all considerably closer to the overall average intensity of 19.60. The 'Face Shape' group was extremely close with an average of 19.65, and the 'Ear's averaged at an intensity reading of 20.21, which was followed by the 'Mouth' group's average of 20.63. The only other facial feature group to be lower than the overall intensity average was the 'Nose' feature group which was averaged at an 18.87 intensity reading. There are few outliers across these four facial feature groups which indicates that they are less important to FaceReader in its interpretation of the 'Surprised' emotional expression. The square-shaped jaws produced slightly higher intensities, but the factor that *FaceReader* focused on was the chin shape. Faces with sharper chins that protrude downwards were found to produce higher 'Surprised' expression intensities, likely due to the perceived further lowered jaw in comparison to the rest of the face. The 'Ears' testing group also had little impact on the intensities of surprised expressions, with the sizes of the ears being unimportant to *FaceReader*, although the ear shapes that faced forwards did produce higher 'Surprised' intensities regardless of the size of the ear.

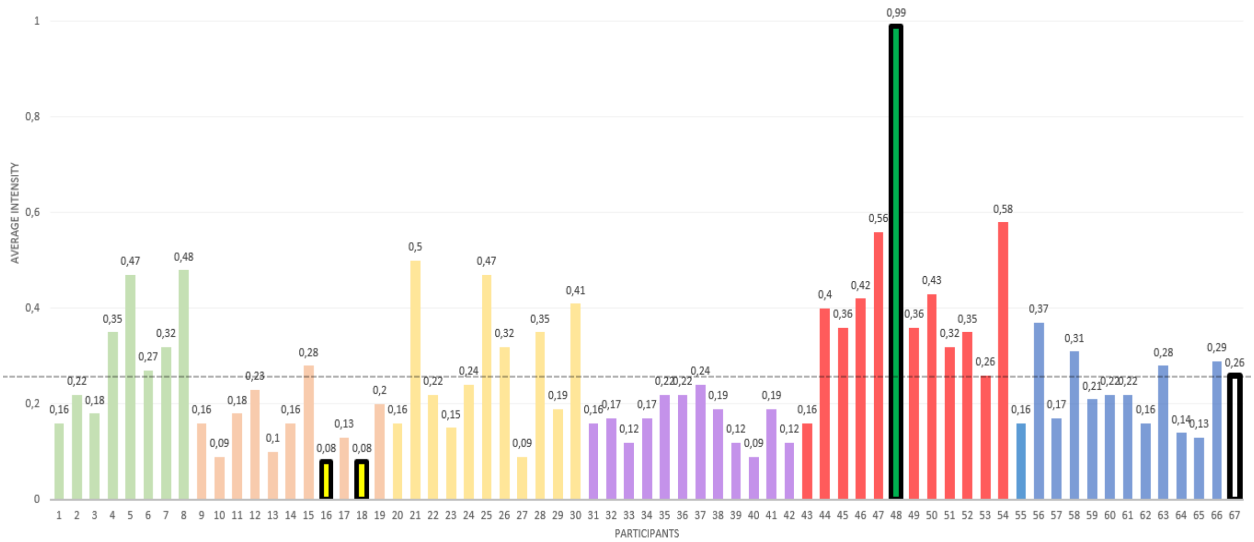
When expressing surprise, the nasal area is supposed to relax as the face widens and the facial features extend apart from one another. The longer nose shapes that extend toward the lips produced lower than average intensity readings. These lower readings were likely due to *FaceReader* interpreting the covered upper lip area as the lower nasal area still being contracted together, whereas the shorter noses provide a clear view of the lower nasal area which is interpreted as the features extending further apart. The 'Mouth' testing group's results were affected by the thickness of the lips. If both or one lip was noted as 'thick', the intensity of surprise rose, and the lowest intensity reading from this group was from a mouth shape that had thin top and bottom lips. *FaceReader* would alter its intensity readings as the thicker lips were interpreted as being more relaxed when compared with mouth shapes with thin lips. While the distance the mouth parts when expressing surprise is important, these readings would suggest that FaceReader places more importance on how contracted the lips and nasal area appear to be to determine the intensity of ‘Surprised’.

5.4.4.2 Surprised Graph - Discussion

What the ‘Surprised’ emotional expression graph brought to focus was that the techniques that FaceReader uses to avoid personal bias does not seem to be effective in all areas of the face. This finding came from how the rounded eyebrows can produce much higher intensities of surprise due to their neutral position being slightly raised and similar to the looked-for movements of surprise. FaceReader can mitigate bias towards sadness in eyes that are already angled to seem sad, and FaceReader is able to remove the ‘Happy’ expression bias that the eyebrow shapes that are angled too far inward would cause. Yet, the bias towards ‘Surprised’ that rounded eyebrows cause, and the bias away from ‘Happy’ and ‘Angry’ caused by small eyes that reduce iris visibility is not accounted for. This problem may not be as persistent when the pre-calibration method is used, but it is a prevalent issue in FaceReader’s continuous calibration use.

5.4.5 Scared Expression Focus

Table 16 Scared Expression Graph Results



5.4.5.1 Scared Graph – Analysis

When analysing the intensity of the ‘Scared’ emotional expression, the most important facials features would supposedly be the ‘Mouth’ and the ‘Eyes’. This focus would be due to intense feelings of fear causing the mouth to open wide, but the lips will be tensed as opposed to the relaxed ‘Surprised’ mouth, and the upper eyelid raise higher to reveal more of the iris. The presented graph does not necessarily follow the traditional ideas of how fear intensities are presented, as both they ‘Eyes’ and ‘Mouth’ groups produce fairly consistent results. Their consistency is shown in their averages, as the overall average is presented as 0.26 and both of these feature groups are almost identical to this value. The ‘Mouth’ group averaged at 0.22, while the ‘Eyes’ group averaged at 0.28. From a glance, the facial features that had the highest impact on *FaceReader’s* analysis of the intensity of ‘Scared’ were the ‘Eyebrows’ and ‘Nose’ feature groups. The eyebrows groups appeared to be consistently lower than average, with an average of 0.15, and the nose features were consistently higher in comparison to the other facial features, with an average of 0.43. The ‘Ears’ were also quite low, averaging at 0.17, and the ‘Face Shape’ group was close to the overall average as well, with an average of 0.31, although these features were less visually apparent in the graph’s results.

The ‘Eyebrows’ feature group produced two identical readings of 0.08 (P16, P18) that were the lowest average intensities across all of the facial features. The shape of these eyebrows was slightly curved, but closer to straight, but they all had a lower inner corner of the eyebrow. This indicates that *FaceReader* interpreted the eyebrows with lower inner corners, the area closest to the nasal, as being less ‘Scared’. The reasoning for this is due to higher intensities of fear involving a raised upper eyelid, so a lowered brow would interfere with how wide the eye would be able to open. Conversely, the highest average intensity reading was found in the ‘Nose’ feature group, which was presented as 0.99 (P48). This reading was notably higher than any other intensity average in this feature group due to there being no other nose shape that is similar to P48. This feature had a large nose-bridge and the largest lower area out of all of the nose shapes. This feature also had a curved tip of the nose, similar to a bird beak, which reached further toward the upper lips which produced the higher intensity reading. This nose shape was ideal for presenting a higher ‘Scared’ intensity as it was able to alter FaceReader’s interpretation of the entire nasal area. The lips seemed more contracted due to the tip of the nose covering the area, creating the illusion of a raised and tensed upper lip.

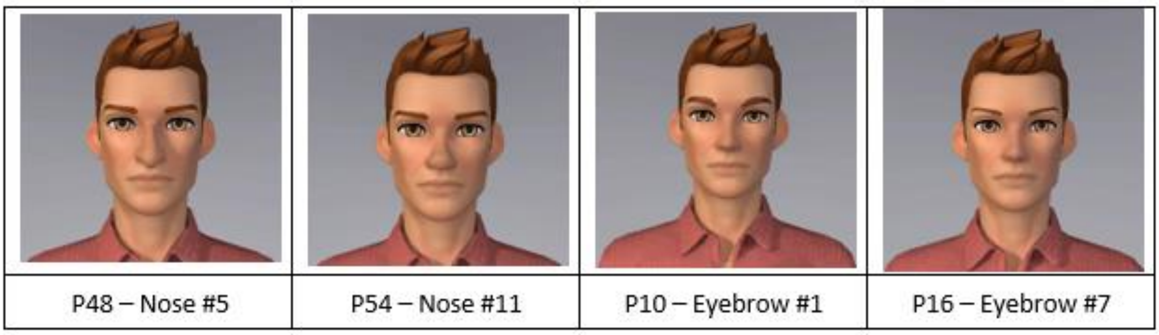


Figure 22 Notable ‘Surprised’ facial features

The ‘Eyes’ group had several fairly high readings that were all presented by eye shapes that had a neutral position that was rounded and more open. The smaller eye shapes that were also characterised by lowered eyelids produced the lowest readings as they lacked the ability to open their eyelids wide enough for *FaceReader* to identify a higher intensity of fear. The ‘Face Shape’ group had higher readings in the feature choices that had rounded jaws and lower intensity readings in the features that had sharper jaw lines.

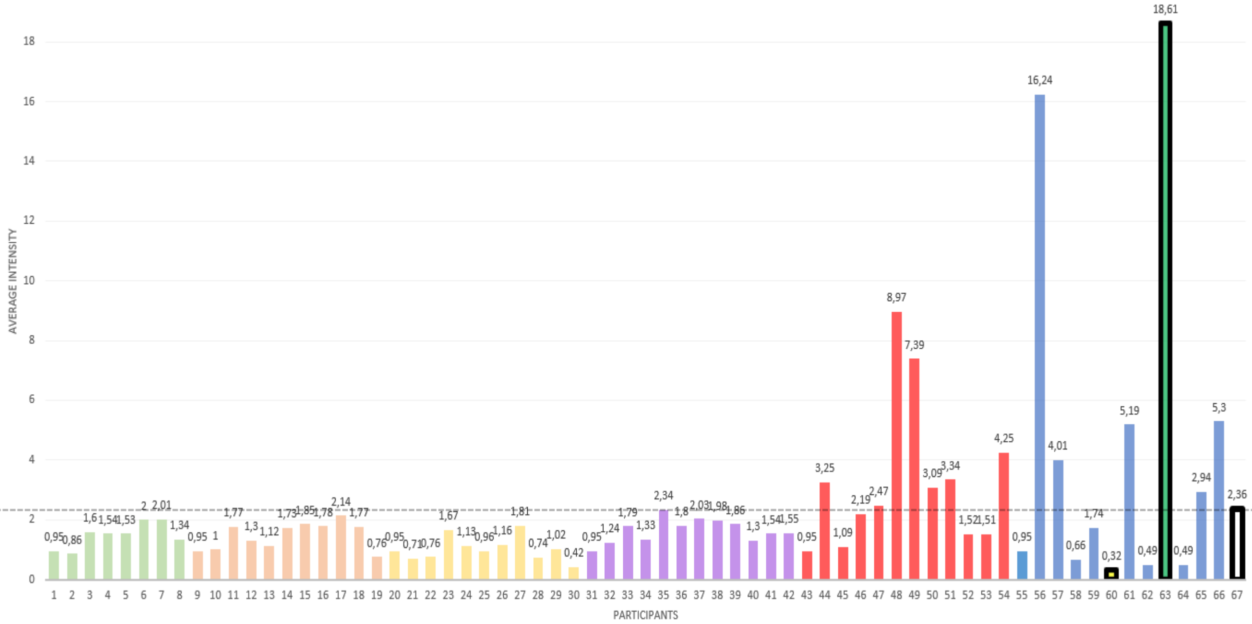
This difference in readings would likely be attributed to the rounded face shapes creating a larger area around the mouth. With a larger area around the mouth, *FaceReader* would interpret there being a higher contraction on the lips even though the movement of the lips will be the same. The different intensities of fear in the ‘Ears’ facial feature group was found to be corelated to the size of the ears, as the larger ears produced slightly higher intensities when compared with the smaller sized ears of the same shape. The reasoning for this would be likely due to the ears affecting the detected edges of the face. With the larger ears, the surface area around the mouth widens and, similarly to the ‘Face Shape’ group, *FaceReader* interprets this as a further contraction of the lips. The final feature group, the ‘Mouth’, had several readings below the overall average due to the intensity of ‘Scared’ relating to the shape of the lips being the area of focus. The feature choices with lips that were square-like in shape produced the highest intensities likely due to *FaceReader* easily identifying their contractions, but there were only four lip shapes that matched this description.

5.4.5.2 Surprised Graph - Discussion

What the ‘Surprised’ emotional expression graph further proved is that *FaceReader’s* interpretation of intensities is heavily reliant on only certain areas of the face for each expression and the way the facial features affect each other is more important than the facial feature itself. As an example, the ‘Face Shape’ group above indicates that the shape of the jaw itself is not important, but the way the jaw shape creates surface area around the mouth is. Another example was found in the ‘Nose’ group, as the length of the nose will alter how FaceReader is able to interpret the contraction and movements of the upper lips. It is necessary to be aware of the way the *Animaze* facial features will alter the overall look of the virtual avatar, as computer vision programs will interpret some of these changes as the expression of emotions. This issue may be due to the ‘continuous calibration’ option, but this claim would have to be tested in a later study.

5.4.6 Disgusted Expression Focus

Table 18 Disgusted Expression Results



5.4.6.1 Disgusted Graph – Analysis

When analysing the intensity of disgust, one would typically focus on the nasal area as when expressing disgust, the face contracts towards its centre. The centre is not necessarily the nose itself, but the area around the nose such as the spaces between the eyes and above the upper lip. From a glance, the readings presented for the intensity of the 'Disgusted' emotional expression were fairly consistent across the first four facial feature groups, but the 'Nose' and 'Mouth' groups produced noticeable outliers due to their influence on the intensity of disgust. The overall average for this emotional expression was found to be 2.36, and each testing group was notably lower or higher in their own averages. The closest group intensity average was calculated to be from the 'Ears', with a value of 1.64, with the other lower intensity averages being the 'Face Shape', 1.48, the 'Eyebrows', 1.47, and the 'Eyes', 1.03. The 'Nose' and 'Mouth' groups had averages higher than the overall average, with the closest being 3.34 from the 'Nose' intensities, followed by 4.75 from the 'Mouth' intensities. These results show an extreme example of how *FaceReader*, and other computer vision programs, would isolate their focus on one area of the face when analysing emotional expressions.

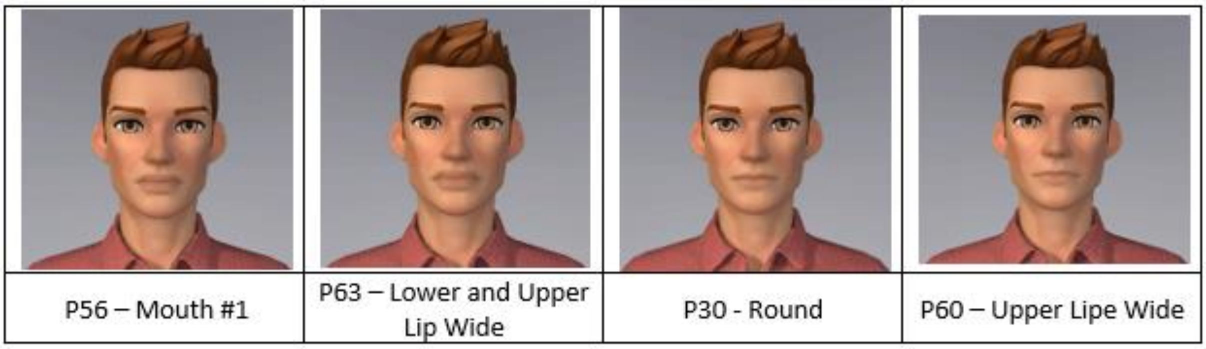


Figure 23 Notable 'Surprised' facial features

The overall highest average intensity readings were both found in the 'Mouth' group as 18.61 (P63) and 16.24 (P56), with the next highest only being 8.97 (P48) found in the 'Nose' group. The higher mouth readings were only apparent in mouths with the thickest lip sizes and were shorter in width. These mouth shapes produced the extremely high intensity readings of 'Disgust' due to the thicker lips being interpreted by *FaceReader* as being contracted higher and closer to the nasal area, due to their size covering more area, and the thick lower lips looked similar to a protruding lip that intense disgust expressions can display. Conversely, the lowest intensity readings were found in the 'Mouth' testing group as well, as the feature choices that had thinner lips (top, bottom, or both) were interpreted as not contracting enough of the face to be a high intensity of disgust. Additionally, the shape of the nose that produced a very high intensity of disgust was the same one that produced a high intensity reading of the 'Scared' emotional expression. The reasoning for this is similar to what was previously discussed, as the longer and larger tip of the nose created the illusion of a higher contraction of the upper lip due to the space it covers.

The 'Ears' testing group produced readings that had few corelations, with the only noteworthy pattern is that the ears should either be extremely large or extremely small to produce higher intensities of disgust, as the lowest readings were all found in the average sized ear participants. The 'Face Shape' intensity readings were slightly affected by the width of the cheeks, but also had a slight focus on the sharper jawlines as they produced higher intensity readings. The significance in the shape of the face was found to be in how the wider cheekbones created an illusion of higher contractions for *FaceReader* due to the lager surface area between the edge of the face and the corners of the mouth. The shape of the eyebrow was not found to be too important, as the 'Eyebrows' intensity readings indicated the higher intensities of disgust came from the angling of the eyebrows. The further down they were angled (inner corners lowered) the higher the expression intensity, likely due to disgusted expressions involving a contraction of the eyebrows towards the middle of the face. Finally, *FaceReader* would use the 'Eyes' to analyse the intensity of disgust through how far the upper eyelid has lowered. The lower average produced by the 'Eyes' testing group was due to there only being three eye shape choices that were both large enough to have the iris identified, and slightly angled downward. The angling downward effect expressed disgust more as FaceReader interprets them as contracting inward and towards the nasal area.

5.4.6.2 Disgusted Graph - Discussion

The 'Disgusted' emotional expression graph provided an extreme example of how FaceReader would isolate certain areas of the face when analysing the intensities of emotional expressions. This can be said as the changes in the main focal areas for expressing disgust in the face, the nose and mouth, noticeably changed the intensity readings produced by *FaceReader*. Additionally, the changes made in the areas not focused on by *FaceReader*, the face shape, eyebrows, eyes, and ears, did not have nearly as extreme an effect on the expression intensities. What these findings also lead one to thinking that not only is the expression of disgust is one of *FaceReader's* most well-trained emotional expressions, as only when there are changes in the focal areas were there different intensities, but also that the *Animaze* virtual avatars have a reliance on the nose to control the other areas of the face. The nose is the focal point, and the shape and size will affect the eye and mouth areas, but the validity of these claims would need to be tested further.

5.5 Expression Focused Graphs – Conclusion

The presented data in this section is the same data from the 'Feature Focused' results section, but they are presented in a bar graph format with every facial feature's (P1-P66) average expression intensity detected by *FaceReader*. By visualising the same data in a different format, new conclusions and theories were found and previously mentioned ideas were further strengthened. A recurring idea that was proposed was that FaceReader's interpretation of the face's expression intensities are based off of one area of the face for each expression. "Happy' will focus on the outer areas of the face, such as the corners of the mouth and the corners of the eyes, while 'Sad' will be highly focused on the shape and size of the eye. 'Angry' was found to be the most evenly corelated feature, with a focus on the whole face but a slightly higher focus on the eyebrow shape, while 'Surprised' had a much higher relation to the shape of the eyebrow. The shape and length of the nose caused a high variation in the intensity of 'Scared', and 'Disgusted' was heavily reliant on the thickness and shape of the mouth. Each facial expression had a noticeable reliance on a certain facial feature, but the more important discovery was finding that how these facial features effect the area around them is more important for *FaceReader* when interpreting expression intensities.

Larger noses would be interpreted as contractions of the nasal area, rounder jaw shapes would be seen as the mouth not moving wider or lower, eye shapes that are angled inwards would create an illusion of the eyebrow moving further over the eye, all of these are examples of how the shape and size of the facial feature would alter the visual clarity of the rest of the face. *FaceReader* would interpret the visual differences as a higher or lower intensity of each expression when that expression's focal point is affected. This is not to say there is a fault in FaceReader's expression recognition or Animaze's virtual avatar, rather that it is important to not only be aware of the facial feature that is being edited but one needs to be aware of how that feature would affect the interpretation of the rest of the face. One would also be able to use this interaction to create an intentional bias towards, or away from, a certain emotional expression. The following section presents an experiment of this assumption, where the presented results are 12 new participants, where P67 – P72 are virtual avatars created to intentionally lean toward certain emotional expressions, and P73 – P78 are virtual avatars created to intentionally lean away from certain emotional expressions. P67 was used previously in this section but was a false participant to represent the overall average of each graph.

5.6 Results: Maximum and Minimum Tests

The presented tables are to be interpreted the same way that the feature focused graphs were read. The rows are the average intensity readings taken from FaceReader and sorted by participants, while the columns are the readings sorted by the emotional expressions. P67-P72 were virtual avatars that were created using the Highest Average Intensity readings from each testing group to choose the facial features that would potentially lean towards a higher intensity of the chosen expression. As an example, P67 was created using the facial features that had the Highest Average Intensities for the 'Happy' emotional expression, P68 was created using the facial features that had the Highest Average Intensities for the 'Sad' emotional expression, and so on. Additionally, P73-P78 were virtual avatars that were created using the Lowest Average Intensity readings from each testing group to choose the facial features that would potentially lean towards a lower intensity of the chosen expression. As an example, P75 was created using the facial features that had the Lowest Average Intensities for the 'Angry' emotional expression, P78 was created using the facial features that had the Lowest Average Intensities for the 'Disgusted' emotional expression, and so on.

5.6.1 Intentional Bias Modelling

Table 22 Highest Results Features

Happy - HAI		**Surprised - HAI**	
P4: Oblong	4,4	P3: Heart	23,55
P10: Eyebrow #1	4,68	P15: Eyebrow #6	25,56
P29: Big Upturn	4,34	P20: Default	21,85
P34: Front Med	5,07	P42: Front XL	22,06
P45: Nose #1	4,29	P43: Default	21,85
P62: Lower Upper Nar	4,55	P56: Mouth #1	22,42
Sad - HAI		**Scared - HAI**	
P8: Round	1,49	P8: Round	0,48
P9: Default	1,32	P15: Eyebrow #6	0,28
P26: Small Monolid	6,38	P21: Almond	0,5
P36: Elf Small	1,97	P37: Scale Small	0,24
P46: Nose #3	2,75	P48: Nose #5	0,99
P55: Default	1,32	P63: Lower Upper Wic	0,37
Angry - HAI		**Disgusted - HAI**	
P1: Default	13,19	P7: Triangle	2,01
P9: Default	13,19	P17: Eyebrow #8	2,14
P20: Default	13,19	P27: Big Monolid	1,81
P31: Default	13,19	P35: Default Small	2,34
P43: Default	13,19	P48: Nose #5	8,97
P55: Default	13,19	P63: Lower Upper Wide	18,61

Table 20 Lowest Results Features

Happy - LAI		**Surprised - LAI**	
P6: Square	2,54	P5: Oval	16,6
P12: Eyebrow #3	3,25	P15: Eyebrow #6	17,7
P25: Upturned	3,2	P26: Small Monoli	9,89
P40: Elf XL	2,61	P37: Scale Small	18,6
P50: Nose #7	2,27	P48: Nose #5	16,2
P58: Upper Narrow	3,25	P62: Lower Upper Na	17
Sad - LAI		**Scared - LAI**	
P7: Triangle	0,67	P1: Default	0,16
P10: Eyebrow #1	0,18	P18: Eyebrow #9	0,08
P29: Big Upturn	0,54	P27: Big Monolid	0,09
P41: Scale XL	0,62	P40: Elf XL	0,09
P49: Nose #6	0,6	P43: Default	0,16
P62: Lower Upper N	0,42	P65: Lower Wide	0,13
Angry - LAI		**Disgusted - LAI**	
P7: Triangle	7,84	P2: Diamond	0,86
P19: Unibrow	3,6	P19: Unibrow	0,76
P26: Small Monol	5,34	P30: Round	0,42
P37: Scale Small	7,31	P31: Default	0,95
P52: Nose #9	8,94	P43: Default	0,95
P63: Lower Upper W	7,42	P60: Upper Wide	0,32

The chosen facial features from Tables 1-6 are laid out in the table below. The green blocks on the left side are the facial features that should, hypothetically, create a virtual avatar that will influence *FaceReader* to present a higher intensity average of each emotional expression. The yellow blocks on the right side are the facial features that should, hypothetically, create a virtual avatar that will influence *FaceReader* to present a lower intensity average for each emotional expression. The aim of this section is to analyse the outcome of using several facial features that by themselves produce higher or lower average intensities of each emotional expression. This will be analysed by following the same testing procedure as laid out in **Section Four** and collecting the new set of data. An outcome that supports the use of specific facial features to produce a higher intensity of an emotional expression would aim to match the following three criteria:

1) The Highest Average Intensity for an emotional expression in the presented table will form a perfect diagonal line from the top left intersection to the bottom right. This would happen as Row 1 (P67) presents the readings from the virtual avatar that uses the highest 'Happy' intensity producing facial features, and column 1 presents all of the 'Happy' intensity readings from the 6 new participants. As the aim is to have the avatars present a higher intensity of a certain emotional expression, in theory P67 should produce the highest 'Happy' intensity reading out of all of these virtual avatars. The same reasoning is true for each participant being named after the emotional expression they aim to produce.

2) The Highest Average Intensity reading from this table should be higher than, or at least close to, the overall Highest Average Intensity from the individual facial feature choices. In theory, by using a culmination of features that produced higher readings on an emotional expression there should be a clear difference between only using one of the presented features.

3) The intensity readings of the other emotional expressions are lower than average, with the hope of them being close to the overall Lowest Average Intensity of the individual facial features. This criterion is not as important, as the intensity of emotions should not have an effect on each other as the amount of time the emotional expressions appear do not change, but previous findings indicate that it is possible for facial features to simultaneously lean toward one emotional expression while leaning away from another.

5.6.2 Increasing Emotions – Summary

Table 24 Results from Emotion Increasing Avatars

Test 7: Max Emotion	Happy	Sad	Angry	Surprised	Scared	Disgusted
P67: Happy-Leaning	4,41	0,11	12,08	20,44	0,05	0,23
P68: Sad-Leaning	3,46	9,53	7,19	10	1,04	1,26
P69: Angry-Leaning	3,79	1,32	13,19	21,85	0,16	0,95
P70: Surprised-Leaning	1,67	0,12	7,35	26,08	0,15	7,27
P71: Scared-Leaning	3,16	0,34	11,45	19,12	1,59	30,63
P72: Digusted-Leaning	2,47	0,28	8,05	19,44	0,56	37,49

The above table presents the findings from the virtual avatars that used the Highest Average Intensity facial features from each emotional expression group. P67 used the features that produced the highest intensity readings of the 'Happy' emotional expression and was able to have *FaceReader* produce the Highest Average Intensity in this table with a reading of 4.41. While higher than the overall average of 3.73, seen in Graph 1, this reading is only the 6th overall highest intensity reading of 'Happy' as there are five individual facial feature changes that produced higher intensities. P68 used the features that produced the highest intensity readings of the 'Sad' emotional expression and was able to have *FaceReader* produce the Highest Average Intensity in this table with a reading of 9.53. This reading is notably higher than any individual feature participant, with the highest being 6.38 (P26).

P69 used the features that produced the highest intensity readings of the ‘Angry’ emotional expression and was able to have *FaceReader* produce the Highest Average Intensity in this table with a reading of 13.19. This was expected as the default avatar design was used for this test, producing the exact same results presented previously. P70 used the features that produced the highest intensity readings of the ‘Surprised’ emotional expression and was able to have *FaceReader* produce the Highest Average Intensity in this table with a reading of 26.08. This was found to be only slightly higher than 25.56, which is the highest reading of the induvial facial feature participants (P19).

P70 used the features that produced the highest intensity readings of the ‘Scared’ emotional expression and was able to have *FaceReader* produce the Highest Average Intensity in this table with a reading of 1.59. While this may not seem much higher than the next highest intensity reading of 0.99 (P48), in relation to the overall intensity averages of this emotional expression (0.26) this is a noticeably higher intensity reading. Finally, P71 used the features that produced the highest intensity readings of the ‘Disgusted’ emotional expression and was able to have *FaceReader* produce the Highest Average Intensity in this table with a reading of 37.49. This was a massive increase in intensity as the highest reading from the individual features were only 18.61 (P63) and the overall average was, in comparison, a very low 2.36.

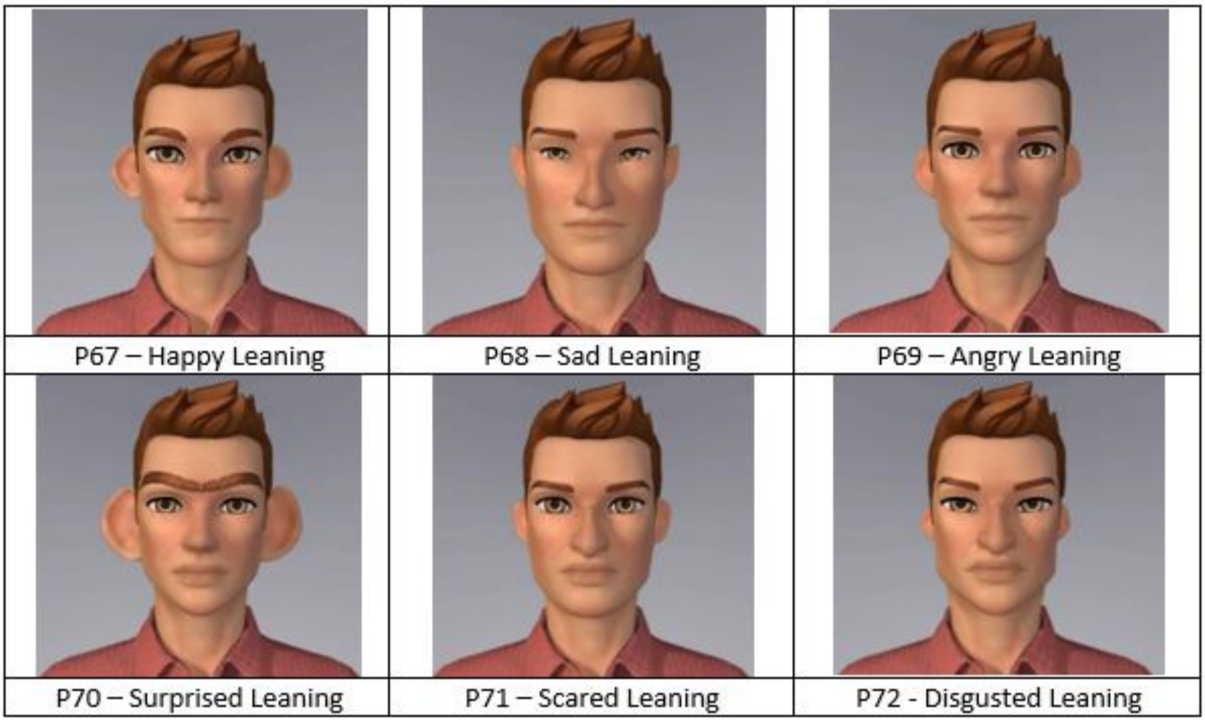


Figure 24 The virtual avatars modelled to produce higher emotional expression readings

5.6.2.1 P67 – Happy Leaning

This virtual avatar was able to produce a fairly high intensity of 'Happy' due to its large face shape and small mouth shape working in tandem to produce a wider looking smile. *FaceReader* would often measure the distance from the bottom of the face to the corners of the lips to see how raised the mouth is, so a smaller mouth and wider jaw is necessary. The upper area of the face is not as important for *FaceReader* when interpreting 'Happy' intensities, but the use of larger eyes and rounded eyebrows would potentially increase the visibility of happier expressions. The same ears that produced the overall Highest Average Intensity was used in this avatar, but the overall average is lower than when this facial feature was the only area changed. It is possible that the sharper jaw and larger eyes were able to influence *FaceReader* to better interpret the edges of the face, as the influence of the ears is not intended to be important for *FaceReader*. While the highest intensity of 'Happy' was not produced, this virtual avatar was successful in reducing the intensities of the other emotional expressions.

5.6.2.2 P68 – Sad Leaning

FaceReader's focal point for the intensity of 'Sad' is placed on the areas around the eyes. The virtual avatar designed to create a higher intensity of this expression has extremely small eyes, and eyebrows that have inner corners that are not too low. These features together alter the visibility of the eyes and create the illusion of a lowered eyelid, a common trait of higher intensity expressions of sadness, and the eyebrows appear to be raised higher as the eyes do not cover much area between them and the eyebrows. The smaller ears would also assist sadness intensities as there would be less interference with the ears and the edges of the face, allowing FaceReader to analyse the positions of the eyebrows more effectively. The shape of the mouth would not have too much of an effect on presence of sadness, but the use of the rounded jaw around the mouth would make FaceReader interpret this virtual avatar as not lowering the corners of the mouth as much and would rather have the mouth appear contracted. While these facial features together were able to produce a higher average intensity of 'Sad' than any of the individual features, they also were able to produce lower than average readings for every other emotional expression which implies a focus on sadness being expressed.

5.6.2.3 P69 – Angry Leaning

There is not much to say on this participant that hasn't been said earlier in this study as the features that created the highest intensity averages of 'Angry' were the default avatar facial features. The evenly spaced-out features, clear difference between feature areas, and the 'average' sized eyes and eyebrows create a virtual avatar that can easily and clearly portray higher intensities of anger. This is due to the entire face being a focal point for *FaceReader* when analysing 'Angry' intensities, so the use of facial features that can clearly indicate movements across the entire face is necessary.

5.6.2.4 P70 – Surprised Leaning

When analysing the intensities of 'Surprise', *FaceReader* was found to focus on the area around the eyes and cheeks. Facial features that are able to portray the eyes widening, cheeks lifting, and the mouth opening are preferable as the more intense the feeling of 'Surprise' becomes the wider the face opens. Through the use of facial features that exhibit these traits, this virtual avatar produced an average intensity of 26.08, which is only slightly higher than the highest individual features average intensity of 25.56. The virtual avatar that was used here had an obscure design that leaned towards larger features. Large, forward-facing ears, a thick lipped mouth, a thick and long unibrow, all of these features were used as when alone they were able to produce high intensities of 'Surprise'. What is worth noting is that these combinations of facial features also were successful in producing lower than average intensities of the emotions of sadness, happiness, anger, and fear, but produced a significantly higher than average reading of disgust. One would assume the reasoning for this is that surprise and disgust share similar traits, but disgust is more apparent when the face contracts towards the nasal area while surprise is more apparent when the face moves out and away from the nasal area. The exact reasoning for the corelation *FaceReader* has for these emotional expressions would require further testing.

5.6.2.5 P71 – Scared Leaning

The 'Scared' emotional expression shares a lot of traits with the 'Surprised' emotional expression, although fear was noted to be more apparent with a contraction in the lower area of the face while surprise is more relaxed around the mouth area. The virtual avatar used to create higher intensities of 'Scared' consisted of a face that would be considered as contracted already. The thick lips but short width of the mouth would be perceived as the lips being tightened together, and the long and pointed nose overflows into the upper mouth area, altering the view of the lower area which would make *FaceReader* interpret this area as being more contracted than normal. Large and wide eyes with slightly thick eyebrows allows there to be a noticeable raise in the eyelids, portraying more fear, and the small ears could be perceived as the face leaning further face and away from what causes the 'Scared' emotional expression. Through these features, this virtual avatar produced a high intensity reading of 1.59, fairly higher than the HAI of 0.99 (P48) found in graph 5. The other emotional expressions were all lower than the overall averages although 'Angry' and 'Disgusted' were higher than average. The average intensity of 'Angry' was calculated as 11.45 where the overall average is 9.96, and the 'Disgusted' emotional expression was extremely higher than average being found to be 30.63 while the overall average was only 2.36. This extremely high reading would be due to the facial features used for this virtual avatar sharing traits with the facial features in the 'Disgusted' avatar (P72). The contraction around the lower nasal area is the most important for FaceReader when interpreting the intensity of fear, and the exact same nose and mouth features were used in both avatars, causing the increased disgust intensity.

5.6.2.6 P72 – Disgusted Leaning

As mentioned above, the lower nasal area between the upper lip and the nose is the most important area of focus for FaceReader when interpreting the intensity of the 'Disgusted' emotional expression. What this virtual avatar (P72) does is actually further reinforces the idea that FaceReader focuses on the areas around the eyes when analysing the intensity of 'Scared'. This is said as P71 and P72 only share the same nose and mouth features, while the eyes, eyebrows, ears, and face shape differ. There are slight changes in the shape and size of the eyes, and the eyebrows are slightly more angled. These changes were enough to produce a higher intensity of disgust while also lowering the intensity of fear. The presented intensity of 'Disgusted' was 37.49, an extremely high reading compared with the overall average of 2.36 found in **Graph 6.** The ability for FaceReader to interpret such an extreme difference in this emotional expression is most useful in further proving that there is a weakness in FaceReader when there are areas of the face that become less visible. When FaceReader is not able to see the area between the nose and upper lip, it assumes a contraction is occurring and interprets this as a higher intensity of 'Disgusted'. What this virtual avatar was also able to do was produce lower than average readings for every other emotional expression, creating a clear example of the possibility of forcing FaceReader to create an artificial personal bias.

5.6.2.7 Table 7: Max Emotions – Analysis

The above results can be used to further identify which areas of the face are focused on when *FaceReader* analyses the intensity of an emotional expression by identifying the facial features used to increase each emotional expression's intensity. Each emotional expression has an area of focus that was laid out throughout **Section 5**, and it was also found that the different shapes and sizes of the facial features available in *Animaze* would affect these areas of the face and how *FaceReader* would be able to interpret them. Table 7 shows that it is possible to use this knowledge to force *FaceReader* to produce higher emotional expression intensity values by using specific facial features that would affect specific areas of the face to promote a certain expression. This sub-section aims to analyse how this was achieved by identifying the facial features used and their intended impact on the face as a whole.

5.6.3 Avoiding Emotions - Summary

Table 26 Results from Emotion Decreasing Avatars

Test 8: Min Emotion	Happy	Sad	Angry	Surprised	Scared	Disgusted
P73: Happy-Avoiding	2,04	0,58	16,81	16	0,24	0,67
P74: Sad-Avoiding	4,56	0,01	12,46	21,03	0,02	1,7
P75: Angry-Avoiding	1,13	3,26	5,42	17,17	0,77	12,75
P76: Surprised-Avoiding	2,56	2,49	5,75	8,19	1,01	10,59
P77: Scared-Avoiding	5,08	0,15	6,3	18,55	0,05	3,1
P78: Digusted-Avoiding	1,9	0,19	4,28	22,44	0,1	0,68

The above table presents the findings from the virtual avatars that used the Lowest Average Intensity facial features from each emotional expression group. P73 used the features that produced the lowest intensity readings of the 'Happy' emotional expression, yet this avatar was not able to have *FaceReader* produce the Highest Average Intensity in this table with a reading of 2.04, while the lowest was 1.13. While not the lowest out of these avatars, the produced intensity average was lower than the Lowest Average Intensity of the individual features value of 2.27 (P50) seen in **Graph 1**. P74 used the features that produced the lowest intensity readings of the 'Sad' emotional expression and was able to have *FaceReader* produce an extremely low average intensity reading of 0.01. not only is this reading notably lower than any individual feature participant, with the lowest in **Graph 2** being 0.18 (P10), this avatar was also almost able to remove the 'Sad' expression entirely.

P75 used the facial features that produced the lowest intensity readings of the 'Angry' emotional expression and was not able to have *FaceReader* produce the Lowest Average Intensity in this table with a reading of only 5.42. This reading is lower than the overall average of 9.96, as seen in **Graph 3**, but not lower than the Lowest Average Intensity that P19 produced (3.6). P76 used the facial features that produced the lowest intensity readings of the 'Surprised' emotional expression and was able to have *FaceReader* produce the Lowest Average Intensity in this table with a reading of 8.19. This was found to be only slightly lower than **Graph 4's** lowest reading of 9.89 (P26) but this virtual avatar was still able to lean the furthest away from higher 'Surprise' intensities.

P77 used the facial features that produced the lowest intensity readings of the 'Scared' emotional expression and was not able to have *FaceReader* produce the Lowest Average

Intensity in this table with a reading of only 0.05. The lowest ‘Scared’ expression intensity was found to be 0.02 and produced by the ‘Sad’ avoiding virtual avatar (P74). While this may not seem much lower than the nearest intensity reading of 0.08 (P16, P18), in relation to the overall intensity averages of this emotional expression (0.26) this is a noticeably lower intensity reading and still lower than all of the induvial facial feature participants. Finally, P78 used the facial features that produced the lowest intensity readings of the ‘Disgusted’ emotional expression and was not able to have *FaceReader* produce the Lowest Average Intensity in this table with a reading of 0.68. Although, the lowest reading was only 0.67 and was found in the ‘Happy’ avoiding virtual avatar (P73). Both of these average intensities are not lower than the Lowest Average Intensity of 0.32 (P60) produced by a single facial feature.

5.6.3.1 Table 8: Min Emotions – Analysis

In a similar fashion to **Table 7’s** analysis, the above results can be used to further identify which areas of the face are focused on when *FaceReader* analyses the intensity of an emotional expression by identifying the facial features that are able to decrease the intensity of each emotional expression’s intensity. **Table 8** shows that it is possible to force *FaceReader* to produce lower emotional expression intensity values for some emotional expressions by merging together the facial features that produced lower expression intensities. This sub-section aims to analyse which of these emotional expression intensities were successfully reduced using an artificial personal bias. This section would also be important in finding out what emotional expressions are seen as more intense when attempting to increase the absence of another expression.

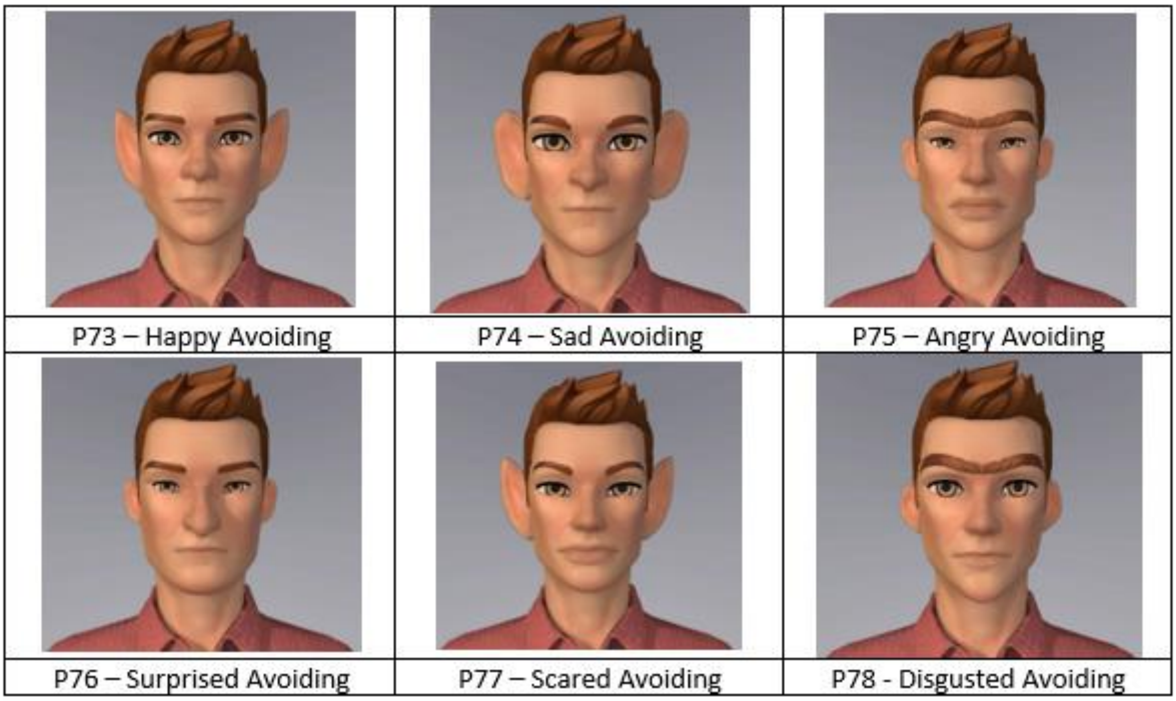


Figure 25 The virtual avatars modelled to produce lower emotional expression readings

5.6.3.2 P73 – Happy Avoiding

As mentioned in the previous sub section, *FaceReader* would aim to analyse the lower face to interpret the intensity of the 'Happy' emotional expression. To achieve a lower intensity of happiness, the focus would be on making the area around the mouth more open for FaceReader to interpret there being less movement. As such, this virtual avatar uses a thin mouth with a short width and a nose that is slightly raised to increase the distance between the upper lip and the nose. This would make it harder to perceive a high rise in the corners of the lips due to the measurement of distance between the nose and mouth. Additionally, the use of a wider jaw and large ears would alter FaceReader's perception of the edge of the face, further reducing the perceived rising of the mouth. While these features were all present in P73, the average intensity of 'Happy' produced was 2.04 which is lower than the lowest individual facial feature's reading of 2.27 (P50) but not the lowest out of these participants. P75 produced the lowest intensity average of happiness with a reading of 1.13, which is less than half of P73. What is worth noting is that this virtual avatar produced an 'Angry' intensity reading, higher than any found in Graph 3, while also producing the lowest "Disgusted' reading out of these participants. Why this is important to note is due to the fact that these facial features were chosen to avoid producing high intensities of the 'Happy' emotional expression, yet these features avoided 'Disgusted' better than the virtual avatar made to avoid that emotional expression.

5.6.3.3 P74 – Sad Avoiding

The eyes and eyebrows were noted to be the most important areas when FaceReader analyses the intensity of the 'Sad' emotional expression. The virtual avatar that aimed to increase the intensity of sadness (P68) was able to do so through smaller eyes that influenced *FaceReader* into believing the intensity of sadness was higher. As such, sadness would be lowered by using larger eyes and eyebrows that would not intrude over the eye area. These traits are seen in P74 as this virtual avatar uses one of the largest eye sizes and eyebrows that are more rounded. Along with these features, a small mouth and higher nose was used to make it harder for *FaceReader* to interpret a contraction of the mouth that is usually seen in higher intensities of sadness. These features together were able to produce a very low average intensity of 'Sad' with a reading of 0.01. This is substantially smaller than the other intensity readings of sadness throughout these experiments, with the closest being 0.15 produced by P77. These features together were also able to produce the lowest average intensity of 'Scared' with a reading of 0.02, highlighting that there are similarities in the facial features that FaceReader uses to analyse sadness and fear. This was one of the more successful tests as this virtual avatar was able to almost remove the intensity of the 'Sad' emotional expression completely through the use of specific facial features.

5.6.3.4 P75 – Angry Avoiding

'Angry' is an emotional expression that *FaceReader* will interpret as being more intense when there is clear visibility of the entire face to analyse the movements between the features. P75 is a virtual avatar that used facial features that would reduce *FaceReader's* ability to interpret anger by using varying shapes and sizes of facial features. The eyebrows used were the unibrow, large and arched eyebrows that are connected and remove the ability to push them together. The eyes are extremely small, the nose is short and thick on the bridge of the nose, and the mouth has thick lips of average width. While these features should result in a low 'Angry' expression intensity, they were only able to produce a reading of 5.42 whereas the lowest reading was 3.6 and came from the unibrow when it was the only changed feature. This was also not the lowest reading out of these participants, as P78 produced a reading of 4.28 while attempting to reduce the intensity of the 'Disgusted' expression. What P75 was successful in doing was it produced not only the lowest 'Happy' intensity average, but also the highest 'Sad' intensity average. While trying to reduce anger, this virtual avatar instead greatly reduced the visibility of happiness which may have been the cause for the rise in sadness intensities. Throughout these tests, 'Angry' seems to be the most unpredictable emotional expression for *FaceReader* to analyse.

5.6.3.5 P76 – Surprised Avoiding

When aiming to reduce the intensity of the 'Surprised' emotional expression, one would aim to reduce the ability of the eyelids and lips to open wide as a more intense expression of surprise is expressed through these areas of the face. P76 was successful in reducing FaceReader's analysis of the intensity of 'Surprised', due to the facial features it presents, as it had an average intensity of 8.19. The next lowest reading was 9.89 (P26), and the overall average in Graph 4 was a much higher 19.60. This was achieved through P76 having the smallest size of eyes and straight eyebrows, making it harder to round the eyebrows out and lift the eyelids higher. Additionally, a large nose that would overflow into the upper mouth area produces a lower intensity of 'Surprised' as the visibility of movement between the upper lip and nose is reduced, making *FaceReader* interpret the lips as not opening wide enough. The face shape and cheeks are also working together to make the virtual avatar seem more contracted together, where a surprised expression is supposed to be more relaxed. This contraction would also be the reason for P76 producing a fairly high intensity reading of 1.01 for the ‘Scared’ expression, which is even higher than the highest individual features intensity of 0.99 (P48). The ‘Sad’ expression intensity was also quite high, 2.49, when compared with the other participants in this group, which further emphasises that the reduction of one expression would often lead to the increase of presence for another emotional expression.

5.6.3.6 P77 – Scared Avoiding

“Scared’ was found to be higher in intensity when the eyelids are able to lift high, there is a contraction around the lower nasal area, the mouth is tense. P77 aimed to avoid higher intensities of fear by using thick and wide lips, which *FaceReader* interprets as less tense, and a short nose that is raised higher, which reduces the visibility of a contracted and tense upper lip by clearly showing the lower nasal area. The eyes used are wide but not tall, as the eyelids already look lower than the other eye choices, which reduces the visibility of the iris and the upper eyelid is not able to rise as much, which is a common trait in intense expressions of fear. The large ears would alter how FaceReader interprets the edge of the eyes and eyebrows, making FaceReader interpret less fear due to the eyebrows not rising and tensing. This virtual avatar was successful in producing a very low intensity of the ‘Scared’ emotional expression, reading at 0.05, but this was not the lowest in this participant group. P74 produced an average intensity reading of 0.02 but the intent of that virtual avatar was to avoid the ‘Sad’ emotional expression, which P77 was able to produce the second lowest average intensity for. These participants shared similar values in every emotional expression except for ‘Angry’, where P74 had almost double the intensity of P77. These findings indicate that the similarities in FaceReader’s analysis of the ‘Sad’ and ‘Scared’ emotional expressions are more connected than originally anticipated. The extent of this connection would need further research.

5.6.3.7 P78 – Disgusted Avoiding

This final virtual avatar in this group was designed to best avoid the ‘Disgusted’ emotional expression. Disgust is heavily reliant on the nasal area, as discussed above, where the shape of the nose and mouth would greatly affect the intensity of this emotional expression. P78 aimed to reduce the intensity of ‘Disgust’ through the use of thin lips and a nose that is short in length. These features would produce high visibility of the lower nasal area, reducing the ability of the virtual avatar to display tension between the nose and lips. Additionally, the rounded and thin bridge of the nose would reduce the ability of this avatar to produce disgust easily as the contraction in the higher nasal area is harder to produce. These features aided P78 in producing a very low intensity of the ‘Disgusted’ emotional expression, 0.68, that was only slightly higher than the lowest reading out of these participants, 0.67, produced by P73 as it aimed to avoid high ‘Happy’ intensity readings. Both of these readings, while low, are not lower than the lowest average intensity of 0.32 produced by P60. What this virtual avatar was successful in was producing the lowest intensity of ‘Angry’ (4.28) while also producing a very high reading of 22.44 for the intensity of ‘Surprised’. These results would again hint to the importance of balance that facial features create, where if a feature is able to reduce the intensity of an emotional expression, it is highly likely that another emotional expression would be increased.

5.7 Conclusion

Section 5.6 presented the findings of an experiment which aimed to test whether or not it was possible to have *FaceReader* provide higher, or lower, emotional expression intensity readings through the use of specific facial features. The results of this test found that this is entirely possible to use certain facial features to influence *FaceReader’s* emotional expression recognition. **Table 7** presented six avatars that were modelled using the facial features that produced the highest intensity averages for each emotional expression in an attempt to have each virtual avatar produce a higher intensity average for that specific emotion. These avatars were all successful in producing notably high expression readings that corelated to their intended emotion. **Table 8** presented six avatars that were modelled using the facial features that produced the lowest intensity averages for each emotional expression in an attempt to have each virtual avatar produce a lower intensity average for that specific emotion. These avatars were also successful in producing low expression readings that corelated to their intended emotion, although the success in recuing emotional expressions was not as great as some readings were higher than the readings of the facial features individually. This test provided value in that virtual avatars can be modelled to intentionally have personal bias. An example of the use of this knowledge would be an animator that wanted to create a character that is pre-disposed to being sad. The use of the above-mentioned facial features would assist the animator is reaching this goal.

SECTION SIX: CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

This section concludes this thesis by restating the research question and the objectives of the study with a discussion surrounding the information presented throughout this thesis. The findings of the study were presented to best answer the research question by comparing the findings in **Section Five** against the literature and information presented in **Section Two** and **Section Three**. This process presented a number of conclusions that validated the contribution this study has made to the knowledge available in this field. Based on the findings presented in this study, this section will also provide the found limitations of this study and recommendations for future studies in this field.

6.2 Reflecting on the Research Aims

The primary aim of the presented research was to answer the following research question:

"How would variations in the facial features of virtual avatars affect facial expression recognition: a study using the Animaze *Library of Avatars with* Face Reader's *computer vision expression detection."*

Therefore, this research project was designed around investigating how the facial expressions detected through *FaceReader* would be affected by changing the facial features of a virtual avatar in *Animaze*.

6.2.1 Guiding Questions

The intent was to use this research project to answer the following research questions:

1. How would changes in the shapes and sizes of key facial features (nose, jaw, eyes, eyebrows, ears, mouth) affect facial expression recognition?
2. Which facial features would be the focal point for *FaceReader* when identifying specific emotional expressions and their intensities?
3. Is it possible to create an avatar in *Animaze* that causes *FaceReader* to produce higher, or lower, emotional expression intensities? If so, how would this be done?

6.2.2 Objectives of the study

Using the above research questions as a guideline, the overall objectives of this study were to:

- Offer insight into the current state of computer vision in the areas of facial recognition, virtual avatars, motion capture technology, and the field of research surrounding these systems to identify a gap in these fields.
- Create an understanding of how facial expression recognition works to better recognise why and how the changing of facial features would affect computer vision.
- Offer insight into interactions between machines by having computer vision programs analyse virtual avatars to potentially identify shortcomings of current technology in these fields.

6.3 Summary of the Study

The presented research aimed to investigate an interaction between machines by using a facial expression recognition program to analyse 3D virtual avatars. Computer vision is becoming a larger part of everyday life as the use of facial recognition spreads to the fields of medicine, security, and education. What is also developing exponentially is the use of virtual avatars as daily life becomes intertwined in the virtual space through social media, video games, and online education. While these fields have been researched thoroughly separately, at the time of this thesis being written there are few noteworthy studies that investigate the interaction between them without the need for human analysis. This is problematic as **Section Two** identified that a major goal in the fields of computer science is the creation of machines that can accurately imitate human emotional expressions AND react to them, yet there is little insight into what would happen when the human mediator between imitation and reaction is removed. The presented research aimed to bridge this gap through a study that had a virtual avatar program (expression imitation) interact with and be analysed by a facial expression recognition program (expression reaction). The intent was to have the virtual avatar present the six basic emotional expressions (Happy, Sad, Angry, Surprised, Scared, Disgusted) while the facial expression recognition program would analyse how prevalent these six emotional expressions are when various facial feature shapes and sizes are used.

The virtual avatar program that was used for the presented research was *Animaze,* a free to use facial motion capture program that gives the user a high level of control over a 3D avatar. *Animaze* allowed me to actively edit the specific facial features on the avatars, namely the: Face

Shape, Eyes, Eyebrows, Nose, Mouth, and Ears, while also providing accurate facial motion capture. As laid out in **Section Four**, this motion capture component was used to have the virtual avatars imitate a 30-second-long video of a real human male acting out the six basic emotions. Once a screen recording of the virtual avatar imitating the video was made, the video of the virtual avatar was presented to the facial expression recognition program for analysis. The program used for this study was *FaceReader*, a commercial program provided by *Noldus Information Technology*. *FaceReader* is able to identify important facial features and track their movements to estimate which emotional expression is being presented by the identified face. The justification behind using *FaceReader* is largely due to the noteworthy research surrounding the program, as presented in **Section Three**, which identified that this program has a high success rate in real time facial expression recognition. While the aim of this study was not to determine the accuracy of the expression recognition program, rather the aim was to identify how the expression readings were affected by varying facial features, it is necessary to use an advanced example of computer vision to best understand the interaction between computer vision and virtual avatars.

The process of analysis was repeated using 66 different facial features available in Animaze, but the exact same input video was used each time to keep the facial features as the only aspect that changes. The use of a single consistent input was done to make the varying facial features the only reason for *FaceReader's* expression intensity analysis to differ. The way FaceReader presented its analysis of the emotional expressions was through expression intensities. For this study, the intensity of the emotional expressions was measured from 0-100, where 0 indicates the emotional expression is not present and 100 indicates there is an extremely prevalent and intense emotional expression present. By identifying the differences in expression readings offered by *FaceReader*, the presented research aimed to identify how variations in each facial feature on a virtual avatar would affect expression recognition. The results presented by FaceReader were organised into two sections for analysis: 'Feature Focus' and 'Expression Focus', where the Feature Focus analysis grouped the readings according to the above-mentioned facial features, while the Expression Focus analysis grouped the readings according to the six basic emotions. The splitting of these results was justified by the aim of this study being to analyse the overall affect that varying facial features would have on facial expression recognition, therefore presenting the results from varying viewpoints would provide this research study with a deeper understanding of the machine interactions.

6.4 Discussion of Findings

The presented research study's findings are presented in full in **Section Five** of this thesis, therefore this section will be presenting only a summary of these findings to promote discussion points, and the results of the presented research were split with the intent of answering the three discussed guiding research questions.

6.4.1 Facial Feature Focus

The first question was: "*How would changes in the shapes and sizes of key facial features (nose, jaw, eyes, eyebrows, ears, mouth) affect facial expression recognition?*". To answer this research question, **Section 5.2** organised the results from *FaceReader* into six testing groups depending on which facial feature was altered, and **Section 5.3** summarised and discussed these results in full. Group 1 focused on the 'Face Shape', which edited the shape of the jaw, chin, and cheekbones, Group 2 focused on the 'Eyebrows', Group 3 focused on the 'Eyes', Group 4 the 'Ears', Group 5 the 'Nose', and Group 6 the 'Mouth'. By organising the results according to focus only on one of the facial features that were altered, comparisons can be made between the average intensities of the different shapes and sizes of each facial feature. By analysing the *FaceReader* intensity averages and focusing on one group of facial features at a time, insight was given into how varying shapes and sizes of facial features on a virtual avatar would affect facial expression recognition.

Section 5 introduced 6 tables of information, where **Tables 1-6** presented the average intensity readings interpreted by the *Facial Expression Recognition* component of *FaceReader* and **Table 7** presented the calculated ranges of these expression intensity averages. The six groups of facial features (face shape, eyebrows, eyes, ears, nose, and mouth) were evaluated on how much of an effect they had on changing the expression intensity of the six basic emotions (happy, sad, angry, surprised, scared, and disgusted). The effect of the facial features in displaying the expression intensity of each emotional expression was measured through comparing the Highest Average Intensities, the Lowest Average Intensities, and the range of the recorded average intensities, where a higher range between the HAI and LAI was interpreted as having a higher importance in *FaceReader's* expression intensity reading.

The 'Eyebrows' were found to be the most important facial feature for *FaceReader's* analysis of the intensity of 'Angry', the 'Ears' were found to be most important for the intensity of the expression 'Happy', the 'Nose' was found to be most important for the intensity of the

‘Disgusted’ expression, and the ‘Eyes’ were found to be most important for the intensities of both the ‘Sad’ and ‘Surprised’ expressions. The only facial feature group that was not corelated with having a large effect on expression intensities was the ‘Face Shape’ feature group, which was expected as the shape of the face is not directly corelated with how *FaceReader* interprets emotional expression intensities. This is not to say the shape of the face is irrelevant in affecting the intensity of emotional expressions, rather that this data is suggesting that the shape of the face’s outline is the least important factor when *FaceReader* is measuring the intensities of the six basic emotions.

Several of the expression intensity results also found that *FaceReader’s* intensity readings are visibly affected by the neutral position of the virtual avatar’s facial features, where the neutral position is how these facial features look when no emotional expression is being displayed, due to the way *FaceReader* calibrates each face to avoid personal biases. This is best seen in an example of the ‘Eyes’ feature group. An eye that is angled downward would have inner corners of the eye lower than the outer corner of the eye, and eyes that are angled upward would have higher inner corners when compared with the outer corners. A downward angled eye will produce lower ‘Angry’ intensity readings, while the upward angled eye will produce lower ‘Sad’ intensity readings. The skewing of results here happens as when the virtual avatar is not displaying any emotional expressions, these eye shapes already look ‘Angry’ or ‘Sad’. *FaceReader* is able to recognise this by reading slight anger or sadness as the new neutral for virtual avatars with those facial features. To clarify, this was found to happen in all of the facial feature testing groups and it is a positive feature as without this compensation there would be a much higher presence of these emotional expressions when in fact there should be no emotional expression present.

Another noteworthy finding was that the sizes and starting position of these facial features played a large role in *FaceReader’s* intensity readings. This would be attributed to both *Animaze* and *FaceReader*, where the finer details regarding the exact causes of these results would require further testing. *Animaze’s* contribution to these changes in results is due to how regardless of the position of the facial feature at the start of the motion capture, the virtual avatar will always attempt to finish the features in the same position, but different shapes and sizes will produce different visibilities. As an example, when portraying ‘Happy’ the intensity of this expression is measured through the distance between the lips and how visible the teeth are. In *Animaze*, a thin-lipped mouth shape will be opened the same amount as a thicker lipped mouth, but due to the size of the lips the thin-lipped mouth shape will produce a higher visibility of the teeth and a perceived wider mouth opening. *FaceReader* would interpret the thin-lipped mouth smile as a higher intensity reading of ‘Happy’ when in reality, the input hasn’t changed, and the features should produce similar results.

This is also not to say *FaceReader* is incorrect in its assessment of emotional expression intensities, rather this analysis highlights that when creating a virtual avatar, one needs to be aware of how the facial features affect one another. A larger nose bridge reduces the visibility of eye contractions, larger ears overlap with the face shape and cheeks, smaller eyes reduce visibility of the eyelids, these are just a few examples mentioned throughout this section of the importance of visibility for *FaceReader’s* expression recognition feature and its analysis of expression intensities. What an analysis of these results also indicated was that it would be possible to intentionally create a personal bias by altering certain facial features, which was tested later in the presented research.

While an analysis of the ranges of the expression intensities can help identify the facial feature groups that had a larger impact on the expression intensities, it is possible that the above analysis may be skewed by an individual outlying feature. An outlying feature in this case would one that created a noticeably higher or lower average intensity of an emotional expression, which increased the range of that facial feature group. To further understand which facial features had a large impact on the specific emotional expression intensities, the following section shifts the analysis focus from the facial features to the emotional expressions.

6.4.2 Emotional Expression Focus

The second question for this research study was: *“Which facial features would be the focal point for FaceReader when identifying specific emotional expressions and their intensities?”*. To answer this question the same data from the ‘Feature Focused’ results section is used but they are presented in a bar graph format with every facial feature’s (P1-P66) average expression intensity detected by *FaceReader*. By visualising the same data in a different format, new conclusions and theories were found and previously mentioned ideas were further strengthened. A recurring idea that was proposed was that *FaceReader’s* interpretation of the face’s expression intensities are based off of one area of the face for each expression. “Happy’ will focus on the outer areas of the face, such as the corners of the mouth and the corners of the eyes, while ‘Sad’ will be highly focused on the shape and size of the eye. ‘Angry’ was found to be the most evenly corelated feature, with a focus on the whole face but a slightly higher focus on the eyebrow shape, while ‘Surprised’ had a much higher relation to the shape of the eyebrow. The shape and length of the nose caused a high variation in the intensity of ‘Scared’, and 'Disgusted' was heavily reliant on the thickness and shape of the mouth. Each facial expression had a noticeable reliance on a certain facial feature, but the more important discovery was finding that how these facial features effect the area around them is more important for *FaceReader* when interpreting expression intensities.

Larger noses would be interpreted as contractions of the nasal area, rounder jaw shapes would be seen as the mouth not moving wider or lower, eye shapes that are angled inwards would create an illusion of the eyebrow moving further over the eye, all of these are examples of how the shape and size of the facial feature would alter the visual clarity of the rest of the face. *FaceReader* would interpret the visual differences as a higher or lower intensity of each expression when that expression's focal point is affected. This is not to say there is a fault in *FaceReader's* expression recognition or *Animaze's* virtual avatar, rather that it is vital to not only be aware of the facial feature that is being edited but one needs to be aware of how that feature would affect the interpretation of emotional expressions as a whole. One would be able to use this concept to potentially answer the final research question as with an understanding of the way each facial feature area affects expression recognition, there is the potential to force a higher, or lower, expression intensity reading.

6.4.3 Intentional Expression Intensities

The final question laid out for this research project was: "*Is it possible to create an avatar in Animaze that causes FaceReader to produce higher, or lower, emotional expression intensities? If so, how would this be done?*". To answer this question, the same results used in the previous sections were used but the Highest Average Intensity and Lowest Average Intensity values were identified and used to create 12 new virtual avatars. The first six avatars were modelled using the facial features that produced the HAI values in an attempt to test if this would produce a higher emotional expression intensity, while the second six avatars were modelled using the facial features that produced the LAI values for each emotional expression to test if this would produce a lower emotional expression intensity. The hypothesis here was that if certain facial features would produce higher or lower expression intensity averages by themselves, merging these features would further increase, or decrease, those expression intensities. By analysing the tables found in **Section 5.6**, it was found that it was possible to force *FaceReader* to produce a higher average intensity for the six basic emotions but the results for forcing *FaceReader* to produce a lower average intensity were varied in success.

The way these virtual avatars were designed can be found fully in **Section 5.6** where there was also a clarification of how these combined virtual facial features interacted with *FaceReader*. One such interaction is that some facial features when alone would produce higher intensity averages of certain emotional expressions when compared with when these features we. An example of this was the “Front Facing (Medium)” (P34) ears produced an average ‘Happy’ intensity of 5.07, yet when those same ears were used with the other facial features that produced high intensity averages of ‘Happy’ the intensity average only reached 4.41. This interaction is noteworthy as the use of the same ears produced lower readings when paired with what would theoretically be facial features that lean more toward expressing the ’Happy’ emotional expression. Another notable interaction was how some emotional expressions were greatly increased when multiple facial features that produced high intensities were used together. The ‘Disgusted’ emotional expression is the best example of this as individually, the highest average intensity from a single facial feature was only 18.61, while the virtual avatar that was designed to create a higher expression intensity of ‘Disgusted’ produced a much higher intensity average of 37.49.

The interaction between *FaceReader* and the virtual avatars that were modelled using the lowest average intensity facial features is valuable in that it showed an attempt to force *FaceReader* to produce lower expression intensity averages being only slightly successful. While the virtual avatars that were modelled to produce very low emotional expression averages were successful in reducing the induvial expressions, what resulted was lower expression intensities being found in the virtual avatars designed to lower another emotional expression. A few of these examples were: The ‘Happy-Avoiding’ avatar producing an extremely low ‘Disgusted’ intensity, the ‘Sad-Avoiding’ avatar producing both the lowest ‘Sad’ and ‘Scared’ intensity average of the whole study, and the ‘Disgusted-Avoiding’ avatar producing a very low ‘Angry’ intensity reading. These interactions indicate that *FaceReader,* and likely other computer vision programs, will analyse areas of the face (lower, middle, and upper areas) when interpreting expression intensities. Therefore, if there is a combination of facial features used on a virtual avatar which all effect a certain area of the face there would likely be unintended changes in the emotional expressions presented by the virtual avatar.

The reasoning for the above interactions and their effects on *FaceReader,* and computer vision as a whole, were explained throughout this research study to the best of my ability. While the guiding research question this section was answered, and it was confirmed that certain facial features **can** alter *FaceReader’s* interpretation of the virtual avatar’s emotional expressions, the understanding of these findings is limited by the scope of this study

6.5 Implications of the Study

The findings in this study implied that *FaceReader's* facial expression recognition component will be greatly affected by the changing of facial features on an *Animaze* virtual avatar. While this study provided evidence only for an interaction between these two programs, these findings can be reasonably applied to other computer vision and virtual avatar interactions. This can be said as, when this research study was conducted, *FaceReader* and *Animaze* are highly advanced examples of software within their respective areas and are able to showcase the current sphere of limitations surrounding computer vision and virtual avatars. According to Zafeiriou (et al., 2015: 2) a goal of computer vision as a whole is to develop machines to a point where they can both react to and imitate real human emotional expressions. Therefore, the findings in the presented research would be useful as a steppingstone in assisting the development towards this goal as society moves into more digital space.

Developments in virtual reality, such as Mark Zuckerberg's 'Metaverese', imply that the world is shifting its developments into areas surrounding virtual avatars. Thus, knowing how these virtual avatars are able to produce emotional expressions would be vital in the advancement of online communication between avatars. The findings of this study indicate that it is entirely possible to use facial expression recognition programs on virtual avatars, but these findings would be better used to begin the development of models that are trained in recognising and reacting to the facial features of virtual avatars. The current development of virtual avatars would also be advanced by the presented research as the findings above indicate the power that slight alterations in the facial features of an avatar can have on computer vision programs. Overall, the presented research found information that one could argue was already proven for human facial recognition, but value was provided in this research being potentially used as a steppingstone towards further development in the use of facial recognition on virtual avatars.

6.6 Limitations and Future Recommendations

Section 6.5 explained that the presented research would be best used in future studies as support towards the development of virtual avatar and computer vision interactions, although there are still other limitations to consider. A recurring issue found throughout the results sections of this research study was the limitation of *FaceReader* due to the current methods of building facial recognition software. The training and building of facial recognition programs were found to use datasets of thousands of images of human faces to teach the program to recognise patterns that corelate between facial feature movements and emotional expressions. Why this limited this research study was due to the relevance of the results from *FaceReader* being questioned due to the program being trained using only human faces (i.e., no virtual avatars) and these faces had a mixture of intended and genuine facial expressions. It is understandable that the training of *FaceReader* does not have any virtual avatar examples to gather information from as the intent of this program was to be used on human faces, and the results of this study could actually be argued to be more valuable due to the consistency of FaceReader's evaluation of virtual faces when no training was provided beforehand to recognise them. The second limitation was referred to by Loijens and Krips (2021: 13) as the "Posed or Genuine" argument. Facial expressions can be different when they are intended (posed) and when they are natural (genuine).

An example of this would be a posed smile just lifts the corners of the mouth, while a genuine smile contracts the eyes as well. This argument becomes more important in the use of virtual avatars, as every expression by a virtual avatar is technically posed and artificial, so the concern becomes whether the emotional expressions produced by avatars are worth studying. One could argue that the line between genuine and posed emotional expressions have always been blurred. Humans would likely not present many emotional expressions when experiencing things alone, such as there being little facial expressions when watching a scary movie alone. This creates questions such as: would facial expressions that only appear in social settings be seen intended or genuine? Is the fact that a facial expression is natural or posed perhaps even relevant? FaceReader, and likely other computer vision programs, do not make the distinction of posed or genuine facial expressions, but I believe this would be an interesting avenue of study to answer the above proposed questions.

I further recommend that future studies in this field are careful to note the exact intention of the study used as this field is growing exponentially every year. The developments in technology and the advancements of computer vision have made the knowledge in older research papers less valuable as they use outdated software or machines. Time would also be a very important limitation in the field of virtual avatars as society becomes more involved in the virtual world, therefore I would recommend studies on virtual avatars be limited to yearly time frames to keep up with current developments.

6.7 Conclusion

The presented research set out to investigate an interaction between machines by using a computer vision-based program to analyse the emotional expressions presented by virtual avatars. The focus on this interaction was due to a noted gap in research found in the literature review done for this study, as there were few other noteworthy studies which investigated a computer vision analysis without human guidance. Studies in this field would use human participants to either interpret the emotional expressions portrayed on virtual avatars, or the human participants would be analysed by a facial recognition program. This study set out to bridge the gap between the two studies by having a facial recognition program analyse and interpret the emotional expressions portrayed by virtual avatars to create a machine-only interaction. This interaction was investigated by designing a research study around how the changing of the facial features on the virtual avatar would affect the emotional expression readings of a computer vision program.

The programs used for this study were chosen due to their advanced design, accessibility, and their features, which were laid out in full in **Section 4.** The facial recognition program chosen was *FaceReader*, a computer vision program that is able to analyse videos of a person's face and estimate the emotional expressions they are presenting, and the virtual avatar library chosen for this study was *Animaze*. *Animaze* is a library that hosts a large selection of virtual avatars that can be controlled using motion capture technology, while also enabling the user to actively edit the chosen avatar's individual facial features. These programs were used extensively in this study as the avatars in *Animaze* were analysed by *FaceReader* to identify which emotional expressions were being expressed by the avatars. The value of this study was provided through Animaze being used to record 68 recordings of avatars imitating the same input video of a man expressing emotions, but each recording had the same avatar with slightly different facial features. The active changing of the virtual avatar's facial features to identify how slight changes in facial features would affect the readings of *FaceReader's* emotional expression recognition component. The hypothesis was that by analysing the different results provided by *FaceReader*, it would be possible to identify how the facial features of virtual avatars would affect computer vision.

The findings of this experiment were compiled and sorted into two different analysis sections, the first grouped the virtual avatars based on which facial feature they had altered while the second presented the findings of all 68 participants but only focused on one emotional expression. The aim of the first analysis was to identify how altering the shape and size of a specific facial feature would affect the expression recognition analysis. The results of this found that the simple changing of the shape and size of only one facial feature would cause *FaceReader* to produce notably different emotional expression readings, and these different facial features would often corelate to a certain emotional expression. The aim of the second analysis group was to further identify which of these facial features are able to directly affect *FaceReader's* identification of each emotional expression. The results of this analysis were the identification of facial feature areas that *FaceReader* focuses on when analysing the emotional expressions of a face. A third analysis was added after the first two analysis group results were identified, which aimed to investigate if it was possible to force *FaceReader* to identify a higher, or lower, presence of each emotional expression and how this would be done. This test was done through using the results of the previous analysis groups to model virtual avatars that would theoretically lean towards or away from each emotional expression. The results of this test showed it is possible to have *FaceReader* interpret higher or lower emotional expressions through using facial features that exhibit a personal bias for each emotional expression.

In conclusion, the findings of the presented research shows that the use of facial expression recognition programs on virtual avatars is fairly different from interactions with human participants due to the amount of control given over the avatar and the different understanding required for machine interactions. This field of study is also fairly new, as the information above was inferred through referring to similar research done on human participants, and it is recommended that further investigations are required to fully understand the reasoning for some of the found interactions. There are cases of facial features (Face Shape, Ears) that should, according to computer vision research, not have a noteworthy effect on expression recognition, yet they were able to quite clearly change *FaceReader's* interpretation of expressions. Additionally, it was found that the facial features themselves were not often the cause of *FaceReader's* different readings but rather the effect the features had on the areas around them were more important. Long noses covering the mouth, eyebrows that cover the upper eye, and thick lips which covered more area of the lower nasal-area, these are a few examples of how the interaction between facial features was often more significant to *FaceReader's* expression analysis as they would alter the overall view of important areas. The value in this study is that it indicates how important it is to be aware of these underlying interactions which only appear through closer analysis.

This concluding section restated the presented research question and the objectives of this research study, and the findings of this study were presented according to the guiding research questions presented in **Section One**. The choices made throughout this research study were aided by the literature review in **Section Two**, which assisted me in arriving at a number of conclusions that contributed to the existing body of knowledge in this field. Based on the findings presented in this study, this section concluded by listing the limitations of the study and recommendations for further research. This study as a whole was compiled with the hope of helping further the field of knowledge surrounding machine interactions, computer vision, and virtual avatars.

Bibliography

AlterCam. (2014). PC [Program]. Bolide® Software**:** Tambov, Russia.

Animaze. (2021). PC [Program]. Holotech Studios Inc.: California.

Bailenson, J. & Blascovich, J. (2004). Avatars. *Encyclopedia of Human-Computer Interaction, Berkshire Publishing Group*, 64-68.

Bartneck, C. (2002). *eMuu – An Embodied Emotional Character for the Ambient Intelligent Home*. PhD thesis: Technische Universiteit Eindhoven. ISBN 9038618875.

Barrett, L. F., & Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science 15 (2)*, 79–83.

Bazrafkan, S., Nedelcu, T., Filipczuk, P., & Corcoran, P.M. (2017). Deep learning for facial expression recognition: A step closer to a smartphone that knows your moods. *2017 IEEE International Conference on Consumer Electronics (ICCE)*, 217-220.

Bernal, G. & Maes, P. (2017). Emotional Beasts: Visually Expressing Emotions through Avatars in VR. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*: ACM, PP 1-8.

Bouchra, N. Aouatif, A., Mohammed, N., and Nabil, H. (2018). Deep Belief Network and Auto-Encoder for Face Classification. *International Journal of Interactive Multimedia: Artificial Intelligence*.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Murray.

Deshmukh, R., Paygude, S., & Jagtap, V. (2017). A Comprehensive Survey on Techniques for Facial Emotion Recognition. *International Journal of Computer Science and Information Security*, 15(3), 219-224. DOI: 10.1109/ICCONS.2017.8250725.

Dores, A. R., Barbosa, F., Queirós, C., Carvalho, I. P., & Griffiths, M. D. (2020). Recognizing Emotions through Facial Expressions: A Largescale Experimental Study. *International Journal of Environmental Research and Public Health 17(7420),* 1-13. DOI: 10.3390/ijerph17207420.

Di Fiore, F., Quax, P., Vanaken, C., Lamotte, W., & Van Reeth, F. (2014). Conveying Emotions through Facially Animated Avatars in Networked Virtual Environments. Conference Paper in: *Lecture Notes in Computer Science, June 2008*. DOI: 10.1007/978-3-540-89220-5_22 · Source: DBLP.

Dupré, D., Krumhuber, E. G., Küster, D., & McKeown, G. (2019). Emotion Recognition from Posed and Spontaneous Dynamic Expressions: Human Observers Versus Machine Analysis. *Emotion*. Advance online publication. DOI: 10.1037/emo0000712.

Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R.C., & Mathiak, K. (2008). Recognition Profile of Emotions in Natural and Virtual Faces. *PLoS ONE 3(11)*: e3628. DOI: 10.1371/journal.pone.0003628.

Elio, R. Hoover, J.Nikolaidis, I. Salavatipour, M. Stewart, L. & Wong. K. 2011. *About computing science research methodology*.

Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest* 8, 151-158.

Ekman, P. & Friesen, W.V. (1975). *Unmasking the Face.* New Jersey, Prentice-Hall Inc.

Ekman, P. & Friesen, W.V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior 6(4)*, 238-252.

Ekman, P. & Friesen, W.V. (1986). A new pancultural facial expression of emotion. *Motivation and emotion, 10(2),* 159-168.

FaceReader. (2022). PC [Program]. Noldus Information Technology: Netherlands.

Faita, C., Vanni, F., Lorenzini, C., Carrozzino, M., Tanca, C., & Bergamasco, M. (2015). Perception of Basic Emotions from Facial Expressions of Dynamic Virtual Avatars. Conference Paper. DOI: 10.1007/978-3-319-22888-4_30.

Fabri, M. (2006). *Emotionally Expressive Avatars for Collaborative Virtual Environments*. PhD thesis, Leeds Metropolitan University, Leeds.

Geraets, C.N.W., Tuente, S. K., Lestestuiver, B.P., van Beilen, M., Nijman S.A., Marsman, J. B. C., & Veling, W. (2021). Virtual reality facial emotion recognition in social environments: An eye-tracking study. *Internet Interventions 25(2021)*, DOI: 10.1016/j.invent.2021.100432.

Garau, M. (2003) *The Impact of Avatar Fidelity on Social Interaction in Virtual Environments*. PhD Thesis: University College London.

Grafsgaard, J.F., Wiggins, J.B., Boyer, K.E., Wiebe, E.N., & Lester, J.C. (2013). 'Automatically recognizing facial expression: Predicting engagement and frustration'. In: *Proceedings of the 6th International Conference on Educational Data Mining*: pp. 43-50.

Gullion, J. S. & Schafer, L. An Overview of Arts-Based Research in Sociology, Anthropology, and Psychology in Leavy, P. 2017. *Handbook of Arts-Based Research*. New York: Guilford Press, pp: 511-525.

Hustadt, U. 2009. *Lecture 6: Research methods*. Research Methods in Computer Science. University of Liverpool: 102-117.

Jiang, J.L. & Loe, K.F. (2003). S-adaboost and pattern detection in complex environment. *In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Madison, WI, USA.

Johnson, E., Hervás, R., de la Franca, C. G. L., Mondéjar, T., Mondéjar, S. F., & Favela, J. (2018). Assessing empathy and managing emotions through interactions with an affective avatar. *Health Informatics Journal 24(2)*,182–193.

Kazemi, V. & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, 1867-1874. DOI: 10.1109/CVPR.2014.241.

Kuo, C-M., Lai, S-H., & Sarkis, M. (2018). A Compact Deep Learning Model for Robust Facial Expression Recognition. Conference Paper: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).* DOI: 10.1109/CVPRW.2018.00286

Li, S. & Deng, W. (2018). Deep facial expression recognition: A survey. *arXiv 2018*. arXiv: 1804.08348.

Li, S.Z. & Zhang, Z. (2004). FloatBoost learning and statistical face detection. *IEEE Trans. Pattern Analysis 26*, 1112–1123.

Loijens, L. & Krips, O. (2021). FaceReader Methodology Note. White Paper. Accessed: September 18, 2021. Available at: www.Noldus.com.

Lozano-Monasor, E., López Bonal, M., Fernández-Caballero, A., & Vigo-Bustos, F. (2014). Facial Expression Recognition from Webcam Based on Active Shape Models and Support Vector Machines. *Lecture Notes in Computer Science 8868*, 147-154. DOI: 10.1007/978-3-319-13105-4_23.

MacDorman, K. F., Green, R. D., Ho, C-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior 25(3)*, 695-710.

Machado, P. P. P., Beutler, L. E., & Greenberg, L. S. (1999). Emotion Recognition in Psychotherapy: Impact of Therapist Level of Experience and Emotional Awareness. *Journal of Clinical Psychology 55(1)*, 39-57.

Magdin, M., Turcani, M., & Hudec, L. (2016). Evaluating the Emotional State of a User Using a Webcam. *International Journal of Interactive Multimedia and Artificial Intelligence 4*, 61-68. DOI: 10.9781/ijimai.2016.4112.

Magdin, M., & Prikler, F. (2017). Real Time Facial Expression Recognition Using Webcam and SDK Affectiva. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(1), 7-15. DOI: 10.9781/ijimai.2017.11.002.

Magdin, M., Benko, L., & Koprda, S. (2019). A Case Study of Facial Emotion Classification Using Affdex. *Sensors 19(2140)*, 1-17. DOI: 10.3390/s19092140.

Magnenat-Thalmann N., Primeau, E., Thalmann D. (1987) Abstract Muscle Action Procedures for Human Face Animation. *The Visual Computer, Springer 3(6).*

Manninen, T. & Kujanpää, T. (2002). 'Non-Verbal Communication Forms in Multi-player Game Sessions', in X. Faulkner, J. Finlay, & F. Détienne (Eds.) *People and Computers XVI – Memorable Yet Invisible*, pp. 383-401. BCS Press, London, ISBN 1-85233-659-5.

Noh, J. & Neumann, U. (1998), A Survey of Facial Modelling and Animation Techniques. *University of Southern California Technical Report*, 99-705.

Nichols, J. A., Chan, H. W. H., & Baker, M. A. B. (2019). Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical Review 11*, 111–118. DOI: 10.1007/s12551-018-0449-9.

Noldus Information Technology. (2020). *Facial Action Coding System (FACS).* White Paper. Accessed: September 18, 2021. Available at: www.Noldus.com.

O'Regan, K. 2003. Emotion and E-Learning. *JALN 7(3)*, 78-92.

Platt, S.M. & Badler, N.I. (1981). Animating facial expression. *ACM SIGGRAPH Conference Proceedings 15(3)*, 245-252.

Park, S., Kim, S. P., & Whang, M. (2021). Individual's Social Perception of Virtual Avatars Embodied with Their Habitual Facial Expressions and Facial Appearance. *Sensors 21(5986)*, 1-18. DOI: 10.3390/s21175986.

Parke, F. I. (1972). Computer generated animation of faces. *In Proceedings of ACM National Conference*, 451–457.

Parke, F. (1982) Parameterized modelling for facial animation. *IEEE Computer Graphics and Applications 2(9)*, 61-68.

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review 9(4)*, 278–311.

Pasquale, F., Percannella, G., & Vento, M. (2014). Graph matching and learning in pattern recognition in the last 10 years. *International Journal of Pattern Recognition and Artificial Intelligence 28.*

Pasquariello, S & Pelachaud, C. (2002). *Greta: A Simple Facial Animation Engine*. DOI: 10.1007/978-1-4471-0123-9_43.

Rajakumari, B. & Senthamarai Selvi, N. (2015). HCI and eye tracking: Emotion recognition using hidden markov model. *International Journal of Computer Science, Engineering and Technology 6*, 90–93.

Ratan, R & Lei, Y. (2021). *What is the metaverse? Two media and information experts explain*. Available at: www.dailymaverick.co.za. [Accessed: 25 Feb, 2022].

Renna, M. E., Quintero, J. M., Fresco, D. M., & Mennin, D. S. (2017). Emotion Regulation Therapy: A Mechanism-Targeted Treatment for Disorders of Distress. *Frontier Psychology*, 8(98). DOI: 10.3389/fpsyg.2017.00098.

Rincon, J.A., Costa, A., Carrascosa, C., Novais, P., & Julian, V. (2019). EMERALD—Exercise Monitoring Emotional Assistant. *Sensors 19* (1953). DOI: 10.3390/s19081953.

Schalk, J., van der; Hawk, S.T., Fischer, A.H., & Doosje, B.J. (2011). Moving faces, looking places: The Amsterdam Dynamic Facial Expressions Set (ADFES). *Emotion 11*, 907-920. DOI: 10.1037/a0023853.

Sakai, T., Nagao, M., & Fujibayashi, S. (1969). Line extraction and pattern detection in a photograph. *Pattern Recognition 1*, 233–248.

Singh, H. 2019. *Practical Machine Learning and Image Processing: For Facial Recognition, Object Detection, and Pattern Recognition Using Python*. New York: Apress.

Slater, M., Howell, J., Steed, A., Pertaub, DP., Garau, M. (2000). 'Acting in Virtual Reality'. in *Proceedings of the third international conference on Collaborative Virtual Environments*, 103-110. San Francisco, California. ISBN 1-58113-303-0,

Solanki, K. & Pittalia, P. (2016). Review of Face Recognition Techniques. *International Journal of Computer Applications 133(12)*, 20-24.

Song, S. L., Shi, W., & Reed, M. 2020. Accurate face rig approximation with deep differential subspace reconstruction. ACM Transactions on Graphics, 39(4): 1-12.

Tian, Y. Kanade, T., & Cohn, J.F. (2000). *Eye-State Action Unit Detection by Gabor Wavelets*. PP. 143–150. Springer: Berlin/Heidelberg.

Tomkins, S. S. (1962). *Affect, imagery, consciousness: Vol 1. Positive affects*. New York, NY: Springer.

Tzimiropoulos, G. 2013. *300 Faces In-the-Wild Challenge (300-W), ICCV 2013* (WEBSITE). Available Online: https: //ibug.doc.ic.ac.uk/resources/300-W/ (Accessed: 25 June 2021).

Van der Vaart, G., van Hoven, B., & Huigen, P.P.P. (2018). Creative and Arts-Based Research Methods in Academic Research. Lessons from a Participatory Research Project in the Netherlands. *Forum: Qualitative Social Research 19(2*), Art. 19. 9, DOI: 10.17169/fqs-19.2.2961.

Wang, H., Song, W., Liu, W., Song, N., Wang, Y. & Pan, H. (2018). A Bayesian Scene-Prior-Based Deep Network Model for Face Verification. *Sensors 18(6)*. DOI: 10.3390/s18061906.

Wang, Y., Lv, Z., & Zheng, Y. (2018). Automatic emotion perception using eye movement information for e-healthcare systems. *Sensors 18(2826).*

Waters, K. (1987). A Muscle Model for Animating Three-Dimensional Facial Expression. *ACM Computer Graphics 21(4).* ACM Press.

Wei, X., Zhu, Z., Yin, L., & Ji, Q. (2004). 'A Real Time Face Tracking and Animation System'. in *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop 5*, 1-78. ISBN 1063-6919.

Wu, X., Yuan, P., Wang, T., Gao, D., & Cai, Y. (2018). Race Classification from Face using Deep Convolutional Neural Networks. *In Proceedings of the 2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM),* Singapore. PP. 1–6.

Yuille, A.L., Hallinan, P.W., Cohen, D.S. (1992). Feature extraction from faces using deformable templates. *International Journal of Computer Vision 8*, 99–111.

Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding 2015(138),* 1-24.

APPENDICES

Appendix A: Ethics Clearance

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)

Registration number: REC-101114-044

17 November 2021

Re: Mr. Keenan van Rooyen (1386104)

Waiver letter number: HRECNMW21/11/10

To whom it may concern,

Mr. van Rooyen is currently a registered Masters student at the School of Arts at the University of the Witwatersrand, Johannesburg. This letter is to confirm that, at the time of writing, Mr. van Rooyen does not need ethical clearance for his Masters study entitled *'Using facial mapping to accurately convey emotions through virtual avatars'.* This decision has been reached based upon a description of the project supplied by Mr. van Rooyen to the University Human Research Ethics Committee (Non-Medical), which has been evaluated by the Chairs and Deputy Chairs. If, however, if Mr. van Rooyen changes the methods of data collection and analysis for this study, this decision may no longer be valid. If such changes take place, this should be communicated to the University Human Research Ethics Committee (Non-Medical) as soon as possible. This waiver letter is valid until 16 November 2024.

Please feel free to contact me should you require any further information.

Thank you.

Yours sincerely,
S Schoeman

Shaun Schoeman (Administrative Officer)
Solomon Mahlangu House, 10th Floor, Room 10004, Jorissen Street, Braamfontein, Johannesburg
Private Bag 3, Wits 2050
T + 27(0)11 717 1408 | **E** Shaun.Schoeman@wits.ac.za | hrec-medical.researchoffice@wits.ac.za
www.wits.ac.za/research/about-our-research/ethics-and-research-integrity/

Appendix B: FaceReader and Animaze Permission

Annet Zander <Annet.Zander@noldus.nl>
to me

Tue, Nov 16, 2021, 1:07 PM

Dear Keenan,

Thank you for your request for a trial version of FaceReader 9.0!

This trial version can only be used by one user at a time and is valid for 14 days after activation!

Please follow these steps to download and activate your new software.

1. You can download the latest version of our software via this link: https://cdn.noldus.com/download/share/software-facereader/pcPeAploLoS0FGS7pdMrVLwTSBVnHV/MTYzNzY2ODc0Nw%3D%3D
2. Install the software on your computer.
3. Your software activation key: BETK1N0Q00G3E8HJ82FM0P9RSC1F15GEDVBE4HG9B6Q5ET.
   You will need this code when activating your new software.
4. Start up the software and choose *to Activate software license key:*
5. Choose to use your software as a:
   1. **Floating license;** you can use your trial license on multiple computers (one at a time), requires an internet connection during use of the application.
   2. **Fixed license;** you can only use your trial license on one computer, does not require an internet connection during use of the application.
5. For floating license usage: repeat step 1 through 5 if you wish to install the software on multiple computers. Please note the software can be used on one computer at a time.
6. Read the quick start guide before you start using the software for your research. For more detailed information, please check out the online help function in the software.

Noldus: FaceReader Inbox ×

Melita Zadel <Melita.Zadel@noldus.nl>
to me, Niek

Tue, Nov 30, 2021, 3:36 PM

Dear Keenan,

Thank you for your email. It was nice to talk with you about your research including FaceReader last Friday.

First of all, my apologize it took a bit longer to get back to you but I had to talk about your request with the product manager of FaceReader, Niek Wilmink (in CC). I got a consent from him that you can use the data from FaceReader also when publishing your research. As discussed in our meeting, we would appreciate if you share your publication with us as soon as it is ready.

During our talk we mostly looked into how different facial features affect the facial expressions analysis. I spoke about this with Niek and he was intrigued by your study as well. He will look into this topic and get in touch with you in the coming weeks.

Please find below a new free trial version of FaceReader (as before, this trial version is valid for 14 days after activation)

You can download the latest version of your software via this link: https://my.noldus.com/download/latest/facereader
Your software activation key: BKRN1-Q0200-GHEBC-J8M5P-0R9A8-SS9TF-I8RZ5UJ. You will need this code when activating your new software.

Allyson Benas <ally@holotech.eu>
to Dragos-Florin, Adina, me

Mon, Nov 8, 2021, 8:18 PM

Hello Keenan,

Your research sounds very interesting! I recommend that you use our newer product, Animaze, for your research. Like FaceRig, Animaze allows people to embody 3D and 2D avatars.

Animaze has a couple of key differences compared to FaceRig:
(1) we built our own powerful render and animation engine
(2) we've updated the features to improve their usability and added new features. Plus, if you ever need help, we have a Discord server and loads of FAQs.
(3) Animaze is free to install and use off of Steam.

Please let me know if you have any questions,
Allyson

On Mon, Nov 8, 2021 at 8:00 AM info@facerig.com <info@facerig.com> wrote:

...

Appendix C: Plagiarism Declaration

University of the Witwatersrand, Johannesburg

School of ....ARTS...........

SENATE PLAGIARISM POLICY

Declaration by Students

I KEENAN VAN ROOYEN (Student number: 1386104) am a student registered for DIGITAL ARTS in the year MA DIGITAL ARTS. I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that ALL the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this in not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: ____________ Date: 2022/03/14

Appendix D: Turnitin Originality Report

Turnitin Originality Report

Processed on: 11-Feb-2022 8:10 AM SAST

ID: 1759878854

Word Count: 44869

Submitted: 1

Masters Thesis – Keenan van Rooyen By Keenan van Rooyen

Similarity Index	Similarity by Source	
5%	Internet Sources:	4%
	Publications:	2%
	Student Papers:	1%