8-2015

Fear Feedback Loop: Creative and Dynamic Fear Experiences Driven by User Emotion

Charlene DiMeglio
cjd8363@rit.edu

Fear Feedback Loop: Creative and Dynamic Fear Experiences Driven by User Emotion

by

Charlene DiMeglio

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

Supervised by

Dr. Joe Geigel

Department of Computer Science

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

August 2015

The thesis "Fear Feedback Loop: Creative and Dynamic Fear Experiences Driven by User Emotion" by Charlene DiMeglio has been examined and approved by the following Examination Committee:

Dr. Joe Geigel
Professor
Thesis Committee Chair

Dr. Arthur Nunes Harwitt
Lecturer

Prof. Phil White
Senior Lecturer

Dedication

To those who cheered me on and went with my crazy schemes.

Acknowledgments

I am grateful for my adviser Joe Geigel for taking me on, and the rest of the committee, Arthur Nunes Harwitt and Phil White, for being supportive and letting me excitedly babble about this.

I am grateful to Iftekhar Naim for guiding me through choosing machine learning algorithms for this project when it was way over my head.

I am grateful to all of my friends and colleagues who lent a helping hand in various ways, making sure I had all the resources I needed to get the job done. Here's to all the teachers, movie providers, computer lenders, rubber duckies, ride givers, and research guides!

I am grateful to each of my subjects for going through my tests. I know it wasn't the most relaxing thing!

Abstract

Fear Feedback Loop: Creative and Dynamic Fear Experiences Driven by User Emotion

Charlene DiMeglio

Supervising Professor: Dr. Joe Geigel

This thesis examines whether it is possible to generate fear-eliciting media that custom fits to the user. The system described uses a genetic algorithm to produce images that get more scary through the generations in reaction to either physiological signals obtained from the user or a user-provided fear rating. The system was able to detect differing levels of fear using a regression trained on EEG and heart rate data gathered while users view clips from horror movies. It was also found to produce images with significantly higher fear ratings at the fifth generation as compared to the first generation. These higher scoring images were found to be unique between subjects.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis intends to expand unexplored area in creative affective computing using a multi-modal approach for emotion detection and a genetic algorithm to power the feedback loop. The system proposed should be able to produce stimuli that will be able to invoke fear in the subject, created by subject's earlier emotional responses. This experiment hopefully will answer whether subjects can power and guide the stimuli to become more intense, and how the results between subjects could differ.

The contributions of the thesis include the genetic algorithm itself. The use of genetic algorithms to classify fear has been done previously, but using a genetic algorithm to produce the fear stimuli in reaction to the detected fear has yet to be done, as far background research has found. This thesis also contributes to methods of personalization of products. Since the experience guided directly by a user's emotional response, the outcome potentially would be very specific to the user. This project adds to research behind affective feedback loops, and the techniques used here can be generalized out for other emotional responses. This work also expands the work that has been done on fear. In prior research, fear was found to be a less-commonly explored emotional response, often appearing with the identified core 6 emotions, but rarely on its own. This work aims to add to the field of fear specific research, with detection and a elicitation playing major roles in the experimental set up.

Affective computing is a budding research area within the computer science community. Being as such, there are many stones left unturned. The bulk of the research done has been in identifying emotions in subjects, and building general models of these emotions.

However, comparatively, very little has been done when it comes to affective computing applications beyond detection. This study proposes a creative feedback loop, in which a regression trained against a user's fear is used to modify images to become more fear-inducing. Fear has been studied less frequently, and thus there is room to add knowledge. It is also a descriptor that is arguably easier to conceptualize. A happy image could be of almost anything, feeding of individual memories, whereas fearful images are more ingrained into our subconsciousness, some learned, some pre-installed through years of evolutionary shaping.

This experiment is relevant to the computer science community in many ways. If successful, computers may be able to react more meaningfully to user emotion. This is important because computers are becoming more and more a part of people's daily lives. Just as preference learning has allowed tasks to become personalized and flow seamlessly through computer-generated suggestion, affect feedback loops can also similarly create smarter applications that custom fit based on information beyond a user's choice. This implementation in particular focuses on optimizing media for a selected emotion. Although it is fear in this experiment, future work can be done to see how well this method applies to other emotions and across media.

1.1 Background

1.1.1 What is Fear?

In order to train a computer to sense fear, it is important to understand what fear is, and how it affects the body. Fear can either be inherited or learned. Inherited fears are those that helped humans survive when the human species was young, and were passed down through genetics [39]. These include the fear of the dark and a fear of snakes. Learned fears come from experience, such as a traumatic life event.

However the fears were acquired, the threat is perceived first by the amygdala, which then passes the message along to the hypothalamus and the midbrain periaqueductal gray

area [39]. These areas in the brain are responsible for physiological and behavioral fear responses, respectively.

We are most interested in those effects we can observe easily and reliably. These include changes in galvanic skin response, pupil dilation, blood pressure elevation, increase in breathing rate, behavior and EEG arousal, and fearful facial expressions [39]. Most of these responses are known to be a part of the flight or fight response, which prepares the body to deal with the threat.

1.1.2 Evoking Fear

Since fear is seen as a negative emotion, one must be prudent to evoke it in an ethical manner. The papers studied all conducted their experiments with care and organization, with varying materials. Emotion elicitation in general has seen many different approaches, from multimedia approaches that incorporate lights, sounds, and visuals [49], to haptic devices that invoke an emotion through touch [46]. The most common approaches seen in the selected studies have either been through images or through films.

For the purpose of invoking fear, films seem to be the method of choice. Images were usually pull from the International Affective Picture System (IAPS), which is sorted by valence and arousal levels. Valence and arousal levels may not translate strictly into different emotional states [30]. For this reason this paper will focus on films as a emotion eliciting material.

Multiple studies [16] [31] [49] [50] have used films as their emotion eliciting material. Films fall flat when used with children or those with mental impairment, since the material must be understood to be most effective [49]. Also, if the subjects attention is thrown to an external disturbance, or if the film has been previously viewed, it could cause the collected data to be skewed. [31] But the extra cognitive and focus power required may be why it has the potential to create strong emotional responses in subjects. It is also seen in a better light ethically than other possibilities for fear elicitation, and tends to be very standardized [31].

One of the challenges for this method would be pin pointing the sought after subjects emotional data from the data gathered. This is an issue because most film clips span much longer than the targeted emotion, and more than one emotion may be drawn from a film. This can be combated by careful editing procedures and by having the subject fill out a post-test survey [31]. It is also helpful to space out emotion-eliciting periods with neutral films to create a controlled baseline for the subject [31]. Examples of films used to elicit fear include The Ring [49], The Shining [31], and The Silence of the Lambs [31].

Faces and Fear

The experimental media consists of an image of a face. The elicitation method was chosen due to the potential to produce the desired affect, having a strong connection with humanity's evolutionary roots. In [36], it is shown that faces hold a viewer attention longer than images of locations. As discussed previously, attention is necessary for proper affective testing. Fear-inducing faces tend to be those that look themselves fearful, or approach the uncanny valley. [33] finds that viewers tend to mirror the emotions of faces viewed. Fearful faces produce a fearful response in a user. A fearful response can also be gained from faces that are slightly inhuman, perhaps holding an expression just out of the bounds of normal human behavior. This face is then said to be approaching the uncanny valley [21].

1.1.3 Detecting Fear

Now that ways of evoking fear have been discussed, gathering the subject's fear response becomes the next main objective. This section will discuss the pros and cons of different ways of recording an emotional response and its relevance to measuring fear. Many of the papers assessed in this survey use more than one method to gather the responses. Multi-modality will be discussed near the end of this section.

Galvanic Skin Response

Galvanic skin response (GSR), or skin conductance, is the change in the skin's levels of electrical conductance as measured by electrodes applied to the skin [20]. Fear is marked by a decrease in skin conductivity, or an increase in GSR [44]. It has been used for emotion detection in [5] [7] [16] [23][24] [34] [50], and for fear specifically in [15] [20] [26] [44] [49].

Most papers viewed in this survey have had positive results using GSR. [23] found it to be a main contributing factor to the classification process ([23] saw a 97.4 percent accuracy rate) and [16] considered it to be the most important factor. This could be because changes in GSR have been linked with changes in the level of arousal [50]. It also has ability to differentiate fear from other negative valence emotions, such as anger or sadness [7]. Not only has it been influential in affective classification, it has been shown to be a comfortable option for subjects in [50]. And, like other methods based on physiological signals, it is not easily faked, as voice or facial expression may be [7].

[49] found that there was noticeable differences in fluctuations between individuals. This could cause inconsistencies in results. However, this issue was resolved by using the first derivative of the GSR waveform, which was an especially good indicator for fear since the changes in the derivative were similar across subjects. Another potential issue pointed out by [23] is that electrode placement may be inconvenient for some applications, since they are best placed on the hand or feet. Also, GSR is susceptible to noise created by the subject's movement. An action like a cough can make noticeable impact on noise levels [23]. [40] did not find that GSR improved emotion discrimination when used in an experiment that also used cardiovascular data.

GSR is a very viable choice for a fear detection system. It has proven influential for classification in multiple studies, as previously discussed. The drawbacks can be worked around, with careful instruction to subjects and signal processing before classification to remove noise.

Fingertip Blood Oxygen Saturation

Fingertip blood oxygen saturation was used in [49]. It is measured using a device clipped on the subject's finger. It was not found to be a significant contributor to the results found in the study. There was no indication of potential benefits for using the method, but since it is clipped to the finger, it may be cumbersome in some applications. In summary, it does not seem like a good approach for fear detection.

Cardiovascular Responses

There are multiple features to track based on cardiovascular responses, and a variety of equipment to gather such data. The most common features found in literature have been heart rate and its variability, often tracked with an electrocardiogram (EKG or ECG), and blood volume pulse, tracked by a photoplethysmography (PPG). ECG uses electrodes in contact with a subject to pick up the changes in cardiac electrical potential over time [6]. PPG uses a device that emits infrared light and records how much is reflected back. It is clipped usually to the finger of a subject [48]. Fear is often associated with a increase in heart rate, but this alone does not differentiate it from other heart rate inducing emotions [6]. ECG was used in emotion detection in [6] [7] [24] [34] [50], and for fear in particular in [15] [26] [40]. PPG was used to measure blood pulse volume in [5] [16] [24] [34] [48], and for fear especially in [15] [20].

As with GSR, multiple studies have found cardiovascular data to be a strong factor in emotion identification. Heart rate was found to be a deciding factor in [49], though they were found to be more complicated waveforms than those from GSR, and thus needed multiple samples. Heart rate variance proved to be a stronger indication of emotion. [40] found that a decrease in variance within a respiration cycle while heart rate was high was a strong indicator of fear. [20] found that large amounts of heart rate acceleration separated fear from happiness, surprise, and disgust, while there was definite differences in heart rate variation between fear, anger, and sadness. Blood pulse volume was indicated as an important deciding factor in [5], and can be used to determine levels of arousal [48].

As pointed out by [6], there tends to be a lot of variation between subjects on how emotions affect physiological states. Also, heart rate is very dependent on multiple factors, such as position of the subject while testing (laying down versus standing for example) [6] and level of activity relative to the subject [24]. Even a small physical movement, such as talking, can have affects on a reading, and standing up could have a great affect on cardiovascular variables than many emotions [24]. [6] found that heart rate played the smallest role in deciding between emotions. [20] pointed out that even though heart rate in itself may not be a clear indicator for emotion, paired with other cardiovascular variables, it can be used. [34] found that heart rate was useful in detecting neutral versus emotive states, while other factors were needed to determine which emotion best described the state.

There are also certain trade offs in the measurement devices themselves. ECG tends to allow for a wide variety of variables to be tracked accurately, but can be rather cumbersome for the subject, as adhesive electrodes must be placed on the body [24]. However, by giving up some accuracy and breath of data, PPG can be used to measure blood pulse variance, which can be used to detect heart rate as well [48]. This gives the benefit of only having to use one sensor, which usually slips onto the subject's finger [24] [48].

Cardiovascular responses can prove to be useful in fear detection. If the experimental settings keep the subject in a consistent level of activity in a similar position throughout the experiment, noise can be minimized. Both PPG and ECG have been shown to be effective methods. If one is trying to create a less intrusive set up, PPG is favorable over ECG, whereas ECG is the method of choice if accuracy and a large number of variable for tracking is desired.

Electroencephalography (EEG)

Electroencephalography (EEG) measures signals that originate from the central nervous system through the scalp [12]. [37] indicated that negative emotions, like fear, are associated with the right frontal region (in charge of withdrawal behaviors) and the prefrontal

hemisphere. This study also shown alpha and beta bands to be important in emotion discrimination. Using this method, neutral states were indicated by equal EEG output from both sides, and negative states from more output from the right half of the brain. This was just one method of the many surveyed in this paper, but it was the one that most directly and clearly indicated how fear may manifest in an EEG reading. Other studies included [12] [17] [34] [38] [41], with those focusing on fear being [8] [18] [22] [37].

[37] indicated that EEG was more resistant to outside factors than other physiological reading, and found a classification rate of 100 percent for fear. [12] found that EEG had better accuracy than peripheral signals on short periods of time. It can also allow subjects a bit more freedom, freeing up hands and allowing subjects to not have to worry about clothing choices [41]. [18] found that Emotive EPOC EEG headsets did not break the test immersion for subjects, although ratings for comfortableness were medium to poor. The same study found that EEG can be a good way to differentiate calm from fear states.

Depending on the type of EEG headset, some components may become tedious to apply and wear. Many require gel, and unless one is seeking readings from the forehead region only, like in [41], subjects will have to deal with the gel getting in their hair. Wiring may also get in the subjects way [41], but wireless models are available and have been used, such as Emotiv EPOC [22]. Other models can give much more information, but can be a bit cumbersome, like the 256-channel device used in [8], which achieved results of 94.88 percent accuracy. However, when compared to other methods that measure central nervous system activity, such as fMRI and PET scans, EEG is a cost effective and noninvasive alternative [17]. [22] also points out that models such as EPOC can track facial expressions, making the tool multi-functional. However, [22] also found it difficult to get clear emotional differentiation from the neurological readings, citing a low signal-to-noise ratio or IAPS images not being a strong enough emotion elicitation method as possible reasons for this result. [34] also mentions the low signal-to-noise ratio, and suggests a great amount of trials to be run to help alleviate this issue.

How a person experiences fear varies person to person, so like with other methods,

this may need to be taken into account if a subject-independent application is sought. [18] found that although the exact responses varied between subjects, they were in some manner comparable.

EEG is another good method for fear detection, although it may be a little cumbersome. It would not be a recommended method for a testing situation with low amounts of trials. It is also can be a bit tedious, especially if the model requires gel and is fully wired. However, if the user is willing to go through the trouble, or can find a model that does not require gel, it seems like EEG is a very good indicator for fear detection, going by what [37] and [8] have found. It is to be noted that [8] used 256-channels, which required many electrodes to be placed on the scalp, which may have lead to the strong classification results, where as many of the other studies mentioned did not have clear or strong results.

Facial Expressions

Despite there being plenty of papers on facial expression detect, very few were found to rely on non-scripted emotions. As [45] mentions, although useful, posed emotions differed from non-scripted emotions but are often used due to the effort needed to elicit a wide variety of emotions versus an actor portraying each emotion. [9] was one of the few studies that evoked emotion from non-actor subjects. They cited that faked smiles do not correlate with other indicators of enjoyment, and that these non-scripted emotions tend to be more informative. Their study focused on enjoyment versus sadness detection, and found sadness to be harder to detect. This does not bode well for fear detection if one compares this to the pattern found in humans. [32] states that disgust and fear are among the emotions that humans had the hardest time identifying, often confusing fear with surprise. [49] also commented that emotions may be be accompanied by a perceived expression, and that so emotions do not have obvious facial manifestations. It was also found that fear and anger were hard to discern from expressions. Beyond this, as mentioned earlier in other sections, facial expressions can be suppressed or altered to a certain extent by the subject, and facial expressions may map to more than one expression, which could cause modeling concerns

[32]. Also, even though using facial expressions can be less intrusive, it may suffer due to lighting and recording resolution [5].

Considering what has been found in literature, it is not recommended to rely on facial expressions alone to identify fear. However, since many implementations do not require sensors to be directly placed onto the user, it could be used in conjunction with other methods with relative ease.

Multi-Modal

Most of the papers surveyed did not rely on only one modality to achieve their results. Often, a mixture of audio-visual and physiological sensors were used, either in fusion, or as a way to narrow down the windows in which emotion occurred.

[5] found that using a combination of both facial and physiological modalities increased emotion recognition results, making ambiguous results clear and strengthening the system against errors. [12] also found similar improvements in results when combining EEG with other physiological readings.

The cons for using multiple modalities has not been fully outlined by many papers. [49] does indicate that using too many measurement tools can be too intrusive to the subject and become impractical. Otherwise, most papers surveyed used two or more methods for data gathering, but did not indicate either way how using the combination had benefited the experiment.

Using multiple methods for data gathering seems to be not only beneficial, but also a common trend in the field. One must choose which modalities fit best for the application, choosing only though that would make a meaningful impact to avoid unnecessary intrusion on the subject.

Summary of Physiological Signal

Fear recognition still has plenty of room to grow. Perhaps because it is a strong negative emotion, there was less focus on it compared to other emotions. However, based on the

papers surveyed, one can come to a reasonable conclusion on how to gather the data for fear recognition. GSR, cardiovascular responses and EEG all pose to as viable options, and research indicates that using a combination can be most effective. For the least amount of intrusion for the most accuracy, a combination of a simple EEG headset, a GSR band, and a PPG would cover most bases. If mobility is desired, there are heart rate monitors and EEG headsets that are wireless that may prove to be useful, but can experience a decrease in quality and variety of information. ECG and EEG's with more electrodes may provide more information at higher resolutions, but can be cumbersome for most user applications.

1.1.4 Feedback Loop

Much of the research in the field of affective computing has been concerned with building accurate models of emotion. Even though the eventual intent for many of these studies is to apply the information gained about the subject's emotional status to alter the state of a program, the focus, in general, seems to have yet to be shifted to this task. However, there are some studies that did implement an affective feedback loop, assessing a user's emotion as a mechanism of change within the program to then in turn affected the user's emotional state. Two of these focused on changing difficulty of gameplay based on affective state, where the other manipulated images.

The two game implementations, [51] and [3], both operated on game difficulty, but with very different intent. [51] moderated the difficulty of a children's arcade game using a measurement of "fun" gained from heart rate, blood volume pulse and skin conductance. Gradient-ascent is used to determine what combination of variables, such as velocity and diversity of positioning of game objects, would create the most fun experience for the individual. In [3], however, the difficulty is increased as a player experiences more fear as relayed in changes in heart rate. The goal is to try to train the player to control fearful reactions to succeed in the game.

Both studies have some sort of predetermined idea of what difficulty means relative to the game, and may be objectively calculated to some extent. However this thesis and

possibly [42] differ in the fact that the resulting direction in which the program may take is not as definite. [42] created an"Affective Mirror" that is similar to a fun-house mirror, except programatically distorts itself based on the amount a user laughs, increasingly so as the laughter also increases. Because the implementation of the changes in the "Affective Mirror" are not discussed in depth, it is unclear whether there were predetermined stages of silliness, or the choices for image distortion were made on the fly. The latter more reflects the case with the experiment of which this thesis is concerned.

1.2 Hypothesis

The aim for this study is to test whether the method proposed can produce images that cause a significant change in the subject's fear levels. This will be assessed through the physiological signals and machine learning algorithms previously discussed. If it is found that fear-inducing images can be produced, then a comparison will be done to see if the images between subjects differ significantly or converge on similar content. This will be done by comparing the genes operated on by the genetic algorithm, given that the same seed image was used. It is this paper's hypothesis that the images relative to their genetic make-up will be significantly different between users.

Once all of the subjects have been through the experiment, the final fear ratings for each subject will be assessed to see if the media samples were able to evoke a significant change per users as well as over all. A comparison will be done between the end state of the chromosomes of those that started with the same seed image to see if these differ significantly.

1.3 Roadmap

The rest of this report is laid out as follows: Chapter 2 begins with a high level overview of the intended system, delving into details and rational in the subsequent subsections. Chapter 3 explains the final implementation and the events that shaped it. Chapter 4 provides

an analysis of the data gathered and the methods used to obtain such. Chapter 5 concludes with the current state of the research, the lessons learned along the way, and future steps to be taken.

Chapter 2

Framework Design

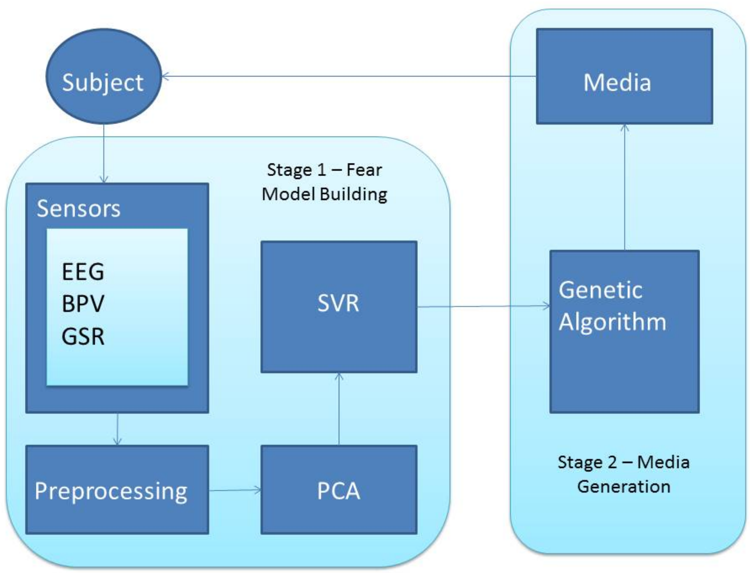


Figure 2.1: An overview of the implementation design.

In order to test the hypothesis, a framework was built to support the fear feedback loop. The solution has two stages associated with it. The first is composed of a series of preliminary trials concerned with training up a model that can accurately assess a subject's fear level through the data collected from the physiological sensors. After this is obtained, the second stage is proceeded with, where subjects are then exposed to the experimental media. Their responses are interpreted by the model specifically trained by their corresponding responses in the previous trials. This numerical rating of their fear level is then fed into a genetic algorithm, which selects those images with the highest ratings to propagate the

next generation of images.

Figure 2.1 gives a high level overview of the framework. As the subject views the video clips during Stage 1, their EEG and BPV signals are measured (GSR, although originally a part of the system was later omitted due to inconsistent functionality of the sensor). The data is then cropped to only include the relevant time spans, and then processed into eight different features per signal (14 for the EEG, 2 for the BPV. Once the features are available from the preprocessing stage, principal component analysis is used to narrow down the 128 features to 10. These 10 features, along with the subjects fear rating of the clip, are then given to train the support vector regression, which outputs a fear rating. During Stage 2, the same steps are taken to gain the 10 features, but the features are used instead as test data rather than training data. The output of the SVR is then used as input to the fitness function of the genetic algorithm. The genetic algorithm produces the media that the subject again reacts to, and that reaction is then fed back into the system through the sensors, completing the feedback loop. sensors, completing the feedback loop.

2.1 Stage 1 - Building the Fear Model

In order to make the system reactive to emotion, it first needed to be able to interpret physiological signals correctly. Subjects were asked to come in for preliminary trials to train the regression to their specific physiological fear responses.

2.1.1 Subjects

The subjects were volunteers gathered from the RIT area. Subject's ages ranged from early 20's to mid 30's. Subjects were briefed on the details of the study and were asked if there were any potential health concerns that may interfere with the study. Subjects also identified which of the movies included in the study they have watched before. Scenes previously unseen are prioritized for viewing per subject. This helped minimize any possible viewing biases.

2.1.2 Materials

The materials gathered to elicit fear from the subjects came from horror movies. These movies were selected from suggestion of lists maintained by bloggers, such as [1] and from previous use in studies such as [49]. A relatively large variety of film clips were chosen in hopes that a variety of fear responses could be gathered from each subject despite unique fears and preferences. A total of 25 clips were gathered from 9 movies. Clips were edited in such a way to give enough context for understanding while maintaining brevity. Clips were shown in such a way to avoid two clips from the same movie from being shown back to back. A consistent chronological order was maintained for all clips such that all viewers would have equal information of the context of each clip despite other clips that may have been shown previously. A full description of the clips can be found in Table 2.1.

Neutral images were shown at the start and in between video clips. These images depicted varied scenery and wildlife.

Table 2.1: A Full Listing of Viewing Materials

ID	Movie	Length (Seconds)	ID	Movie	Length (Seconds)
Alien	Alien	133	Aud1	Audition	28
Aud3	Audition	84	Aud4	Audition	83
Aud5	Audition	59	Aud6	Audition	48
Exo1	The Exorcist	32	Exo2	The Exorcist	39
Exo3	The Exorcist	17	Exo4	The Exorcist	9
Exo5	The Exorcist	41	Exo6	The Exorcist	18
Om1	The Omen	57	Om2	The Omen	71
Ring	The Ring	74	Seven	Se7en	62
Shine1	The Shining	45	Shine2	The Shining	60
Tex1	The Texas Chainsaw Massacre	50	Tex2	The Texas Chainsaw Massacre	65
Tex3	The Texas Chainsaw Massacre	44	Tex4	The Texas Chainsaw Massacre	40
Thing1	The Thing	74	Thing2	The Thing	112
Thing3	The Thing	36			

2.1.3 Hardware

The hardware selected for the study include a Nonin pulse oximeter, a Emotiv EPOC EEG headset, and a Affectiva Qsensor GSR sensor. These were selected due to availability, comfort to subjects, and cost. The corresponding propriety software was used to gather each of the signals. This included, respectively, Sp02 Assistant, TestBench, and Q (although signal

is gathered on the device, it can be viewed through this software). In the end, much of the data gathered by the Q sensor was omitted in the study due to inconsistent success with obtaining data from the sensor.

2.1.4 Experimental Procedure

After volunteers have given consent and have been cleared of any abnormalities that may interfere with the study, the sensors are applied and calibrated. The Q Sensor is the first to be applied since it takes ten minutes to properly warm up. The EEG headset is applied next, moistening the sensors with contact solution before positioning on the head. The pulse oximeter is then put on the subject's right middle finger. The full set up is pictured in 2.2.

Once signals are being measured clearly, the lights in the room are dimmed and the neutral emotion videos, one depicting scenery, are played to gather a baseline. After this is obtained, the user is informed that a clip from a horror movie begins to play. After each fear eliciting film, the subject is asked to rate the level of fear felt on a scale from 0 to 5, with zero being the lowest (not afraid at all), and 5 being the highest (very afraid). The subject then views an image of scenery to return back to baseline levels before moving on to the next film.

2.2 Stage 2 - Generating Media

The second part of the experiment uses the findings found in the first part of the experiment to guide the generation of results.This entails processing the physiological data before feeding it into PCA and the SVR (see Implementation). Once the regression is properly trained, it is ready to receive the data from this part of the experiment.

2.2.1 Materials

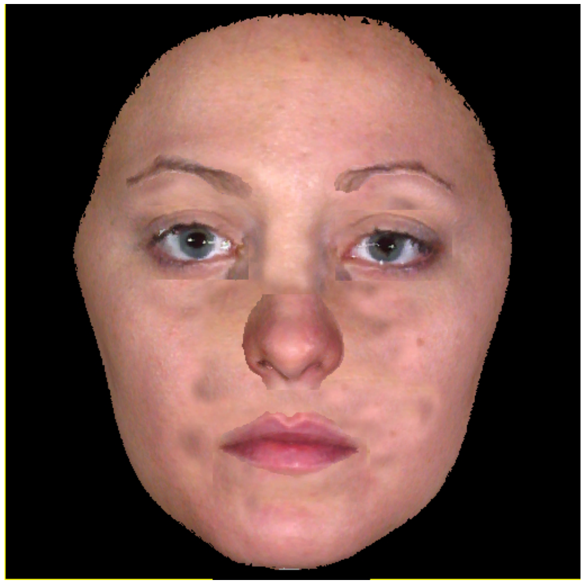
The baseline images used in this part of the experiment are the same as those used in the first part. The emotional media comes from procedurally editing a neutral face from the Department of Computer Science, State University of New York and Binghamton's collection [52]. The face is scrubbed of its prominent features, which later get reapplied procedurally. The base face is shown in Figure 4.6.

2.2.2 Experimental Procedure

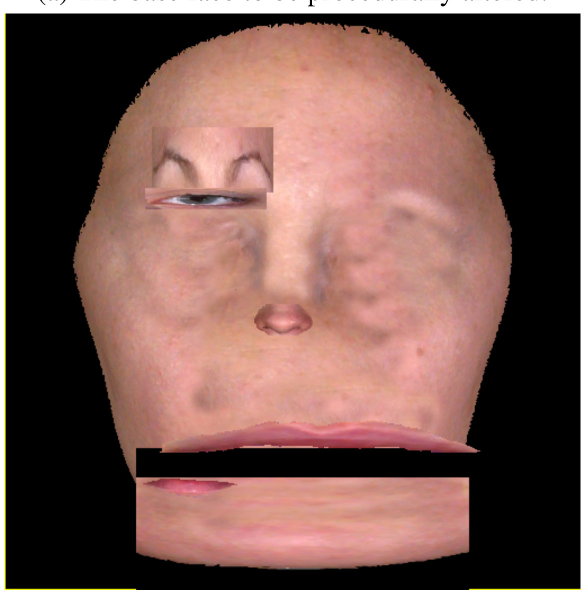
As before, the modalities are applied and calibrated before beginning the study. The lights are then dimmed and a baseline reading is taken for 30 seconds. The subject then views a generation of images. Each generation consists of ten procedurally produced images, spaced between 1-3 neutral images. The pattern is randomly generated and is described in Table 2.2. The first generation for each subject is made up of the same nine randomly generated faces and the face shown in Figure 2.3a. Subjects either passively view the material or are asked the rate the generated images. In between each generation there is a processing period that lasts around 5 minutes. Most of this time is spent lining up the data to correspond to the correct time frames. The next generation of images is also produced during this time. The subjects are requested to participate through at least the fifth generation, if not through the tenth. After the final generation is produced, the subject is given an exit survey and debriefed.

Table 2.2: Viewing Pattern of Experimental Materials. N indicates neutral images.

Seconds	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5	Gen 6	Gen 7	Gen 8	Gen 9	Gen 10
0	N	N	N	N	N	N	N	N	N	N
10	N	N	Image 3	N	N	Image 2	N	N	Image 2	N
20	N	Image 4	N	N	Image 7	N	N	Image 3	N	N
30	Image 3	N	N	Image 10	N	N	Image 5	N	N	Image 9
40	N	Image 6	N	N	N	N	N	N	N	N
50	Image 4	N	Image 2	N	Image 2	Image 3	Image 10	N	Image 8	Image 9
60	N	N	N	N	N	N	N	Image 2	N	N
70	N	Image 8	N	Image 1	Image 8	Image 7	N	N	N	N
80	N	N	N	N	N	N	Image 8	Image 5	Image 5	Image 10
90	Image 10	N	Image 9	N	Image 6	N	N	N	N	N
100	N	Image 7	N	Image 8	N	Image 10	Image 1	N	Image 3	Image 1
110	N	N	N	N	N	N	N	N	N	N
120	Image1	N	N	Image 7	Image 10	N	N	Image 6	N	N
130	N	Image 3	Image 7	N	N	Image 5	Image 7	N	N	N
140	N	N	N	N	Image 1	N	N	N	Image 1	Image 5
150	N	N	Image 4	N	N	Image 9	N	Image 8	N	N
160	Image 7	Image 5	N	Image 2	N	N	N	N	image 7	Image 4
170	N	N	Image 6	N	Image 3	Image 1	Image 4	N	N	N
180	Image 2	Image 2	N	N	N	N	N	N	Image 10	Image 3
190	N	N	N	N	N	N	N	Image 7	N	N
200	N	Image 9	Image 5	Image 6	N	Image 8	N	N	N	N
210	Image 5	N	N	N	Image 5	N	Image 9	N	N	N
220	N	N	Image 10	N	N	Image 6	N	N	Image 9	Image 7
230	N	N	N	Image 3	N	N	Image 2	Image 4	N	N
240	N	Image 1	N	N	N	N	N	N	Image 6	N
250	Image 6	N	N	N	Image 4	N	N	Image 10	N	Image 2
260	N	N	Image 8	Image 9	N	Image 4	Image 3	N	Image 4	N
270	N	Image 10	N	N	N	-	N	Image 9	-	Image 8
280	N	-	N	N	N	-	N	N	-	-
290	Image 9	-	Image 1	Image 5	Image 9	-	N	N	-	-
300	N	-	-	N	-	-	Image 6	N	-	-
310	N	-	-	Image 4	-	-	-	Image 1	-	-
320	Image 8	-	-	-	-	-	-	-	-	-



(a) The base face to be procedurally altered.



(b) Procedurally altered face.

Figure 2.3: The original face shown in (a) has had its features wiped off then reapplied. The face shown in (b) has had said features altered or removed before application.

Chapter 3

System Implementation

3.1 Infrastructure

Originally, Open SSI [47], an open source software for human signal processing written in C++, was to be used. It was chosen because it provides a modularized infrastructure to allow for highly customizable project layouts using XML to interface between components, such as a variety of sensors, camera, and machine learning algorithms. However, it only supported one of the three sensors intended for the study. This is why a change was made to use Attention Tool [2] instead. It supported two of the sensors, EEG and GSR, and has found success in other student projects. However, the resource ran out before the completion of the study, and keeping the data consistent temporally became a issue that needed to be handled directly.

As mentioned previously, the propriety software for the sensors were used, the outputs of which were edited by hand for the first half of the experiment, and by a script in the second half. In the first half of the experiment, it was easier to find the subject-indicated areas in the files directly rather than writing a script to handle the data because the exact points that needed to be extracted varied person to person. In most cases, 15 seconds of data was collected for each clip, 5 seconds before and the 10 seconds after. This was done to be able to capture any sudden changes or general trends indicative of fear. However, in some clips, it was not possible to capture the full 15 seconds, so it was ensured that at least 5 seconds of data would be collected, as it seems studies such as [49] disregard data shorter than 5 seconds in duration. The second half of the experiment is a bit more predictable

in that it is known exactly when a fear signal is expected, and can be anticipated through hard-coding extraction programs. These include one for the pulse oximeter data and one for EEG data (GSR did not end up being included in the study, as will be explained in later sections). These extraction programs, eegsplit.py and hrsplit.py, are throw away programs hardcoded to fit the needs of the study.

3.2 Analysis of Machine Learning Techniques

When selecting a machine learning algorithm for one‘s experiment, it may make sense to use the algorithms used in previous papers as a guide. However, as pointed out by [29], due to the large amount of variation between affective computing experiments, it is not feasible to compare the results of these studies, and due to the lack of standardization, a best practice algorithm has yet to surface. [29] indicates that instead one should focus on the influencing factors of the experiment at hand to choose feature selection algorithms and classifiers. Among these factors are the type and number of features, the number of training vectors available, representation of knowledge, and the adaptiveness of the desired system. The ideal for affective computing according to [29] is to have a system that can continuously deal and grow stronger with a large number of diverse data points.

Regardless of feature selection methods and classifiers, the literature did reveal some useful resources, such as common features extracted and tools useful to machine learning. For instance, [7] indicated that they used PRTools, a pattern recognition library for MatLab, for their study. [20] also proved to be a good resource, giving the formulas for each equation they had used for their analysis. These formulas were common between many of the studies researched, and were referred to as the Picard parameters. These included the mean, standard deviation, mean of absolute values of first differences, means of normalized absolute values of first differences, means of absolute values second differences, and the means of the normalized absolute values of the second differences. This is similar to the values used in [5], [48], [7] and [18]. As an addition to this, the ratio between the minimum and maximum values, the Euclidian distance from the baseline, and the mean

and the standard deviation of the spectral coherence function were also used. The use of the difference between emotive data and the baseline was also seen in [25]. In [48], it was noted that a small statistical feature size was used because their focus rested on the uniqueness of physiological signals between subjects and environments. Since this experiment is user-specific, a small feature size may suit well. Multiple studies [16], [43] that they used normalized values to minimize the effect of subject dependency. However, since the experiment is subject-dependent, normalization does not seem to fit well in this study. Based on this, the raw data is not normalized before computing the values indicated in [20] per modality.

Before describing each of the options for feature selection and classifiers, it is important to revisit the factors that go into choosing such as outlined by [29] relative to this thesis's experiment. The features that are used are those specified in the previous paragraph for each PPM, EEG, and GSR. That makes a total of the number of features to be at least 30. The training vectors come from the data obtained in the initial part of the study. In that portion, the subject views fear-eliciting material while physiological signals are obtained. After each segment, the subject rates their responses. This is be used to judge the effectiveness of the machine learning algorithm. The output to the algorithm should be a score, ranging from 0 to 5, with 5 being the most afraid and 0 being not at all. The system should be able to adapt to an individual user, and thus does not need to full adaptive power a multi-user system would need.

Because having many features may increase the risk of overfitting [7], feature selection is necessary. The features were commonly selected in literature via principle component analysis (PCA) or sequential forward selection. PCA, or principle component analysis, was deemed to be the most suited to affective data feature selection when compared to mutual information and concatenation [5]. However, the success of PCA stands on a large number of independent variables [40] that are similarly scaled. It is uncertain that the full set of data points will be completely independent or dependent, since responses differ from person to person and thus the amount of dependence between physiological signals may

also vary. The scaling of each piece of data may also differ drastically. Since the intended system relies on variables whose relationships are unclear, it is undecided whether PCA will be applicable to this problem. If PCA proves inconclusive, robust PCA may be tried. Sequential forward selection is a greedy approach and does not guarantee optimal results. This study uses PCA to select features, a choice that will be re-evaluated after initial testing.

In literature, many different classifiers were used distinguish different emotive states. Of these, support vector machines (SVMs) appeared the most frequently, appearing in [25], using a Gaussian radial basis function kernel, [5], using libSVM to create a classifier that used a radial basis function kernel, and [20], which found that a linear kernel works best for their study. SVM may not fit well in this study, due to the need for a value as output versus a category. This difficulty is also present in nearest k-neighbors, used in [7]. SVMs also require parameters to be carefully chosen, which should be handled by our feature selection algorithm previously discussed [7]. Another option is to use support vector regression (SVR). SVR is very similar to SVM, but because it is a regression, the output would be a value versus a category. We will be using a SVR as one of the potential classification schemes in preliminary testing.

Other methods were presented in the literature, but they did not seem as promising. A genetic algorithm was used in [51] to minimize the difference between detected and reported responses. Genetic algorithms however would not be best suited for this portion of the study due to the low amount of trails available. This is a similar issue with feedforward neural networks, used by [48]. There is not enough data to properly train the network without the danger of overfitting.

3.2.1 Machine Learning Implementation

Features are drawn from the EDA reading of the GSR sensor, each node of the EEG, and both the oxygen levels and heart rate from the pulse oximeter, which are stored in plain text files. The features include the Picard parameters as described in [20], as well as the

max/min ratio and Euclidean distance parameters described in the same paper. These features per sample are stored into one file that gets passed on to the machine learning program. The features are selected using principle component analysis (PCA) and classified using support vector regression (SVR) as discussed in the previously. From the 128 features collected, PCA outputs ten. The SVR is initialized to the default parameters as given in the scikit-learn Python libraries [4], which are used to formulate both the PCA and SVR processes. The model is then validated using leave-one-out validation. Answers are checked as to their relative correctness, and the kernel is tweaked to obtain satisfactory results.

This implementation is feasible in the fact that it produces the results required, but when needed to predict the fear values from viewing the experimental media, it is a bit more cumbersome than intended. There is quite a bit of work needed to be done by hand before starting the second part of the experiment due to lack of synchronization between the modalities. During the experiment, the process runs a bit more smoothly due to its predictability, but there is still a small bit of overhead when it comes to correctly formatting the data for the system, and then displaying the resulting images in the described method.

3.3 Genetic Algorithm

An interactive genetic algorithm is used to create the media for the affective feedback loop portion of the study. Genetic algorithms optimize content in respect to a selected aspect. Here, we optimize the media to become more fear-eliciting. Much of the research explored in relevance to affective computing used genetic algorithms to classify emotion felt by a subject. Here, however, we use the classified emotion as an input to the system, and the media as the output.

As far as images are concerned, it seems as though images tend to be more common input than affect. In [11], images of lips are inputs to a genetic algorithm used to classify the the emotion portrayed by such lips. In [10], a genetic algorithm uses vector graphic polygons to converge on a given target image using pixel color to guide the progression. The desired effect is not duplication, but rather stylization. This is closer to this study's

intent than [11], in the fact that it is a creative process versus a classification one. Where it differs, though, is that a base image is provided, and the progression is away from the given image based on subject emotion. [19] provides a good basis of understanding for manipulating 3D facial images via genetic algorithm. The exact make-up of the chromosomes for the experimental media will be discussed in a subsequent section, however this paper is important to this study in the fact that it shows a facial image as output of a genetic algorithm. The intent of [19] is to expand the population of facial images. This is very similar to the need for genetic algorithms in this study: to provide a diverse set of stimuli in hopes of convergence on fear-eliciting content. The driver of the algorithm for [19] was diversity, where as diversity too must be considered in the fitness function along side the subject's emotion.

The fitness function is the detected amount of fear caused by said media. As previously stated, this floating point number ranges from 0, being the least fit, to 5 being the most possible fit, however this number should be capped to the highest rating achieved while viewing the movie clips. For instance, if a subject rated the scariest clip in the set as a 3, it should be expected that the scores for the experimental media range from 0 to no higher than 3, since extrapolation can be quite error prone.

The first generation of images are be seeded with the same randomly generated set per subject. After the physiological data gathered while viewing each image has been processed and checked against the regression, the fittest four individuals out of ten are then produce nine children. Each of the four individuals are crossed to produce a child with one another, totaling six children images from this process. The highest scoring image is then again crossed with each of the other three fittest individuals, producing three more children. The last child is randomly produced to prevent premature convergence at a local maximum. In a second incarnation of the genetic algorithm, only the first six children are produced, and four random children are added in. The algorithm was later altered in this way to provide more genetic variety that seemed to be lacking in the first version. A third version of the algorithm was employed that took the top four uniquely scoring individuals.

This provided genetic variation that would not otherwise be seen.

The children go through a mutation stage before joining the newest generation. The mutation stage has ten percent chance of occurring, with each of the four different mutations having equal chance of occurring. The four mutations include duplication of a whole facial feature, the addition of an extra bit in the encoding of the face (see Chromosomes for more information), the removal of a whole facial feature, and the removal of a random bit in the encoding of the face. After each face has been validated such that each of the facial features fits within bounds of the image, the next generation is ready to be rendered. The highest scoring individual per generation is noted, and the highest per person is kept as outputs to the experiment. The core processes of the genetic algorithm are kept in the Genetic.java file. An top-level explanation of the process is shown below in Figure 3.1.

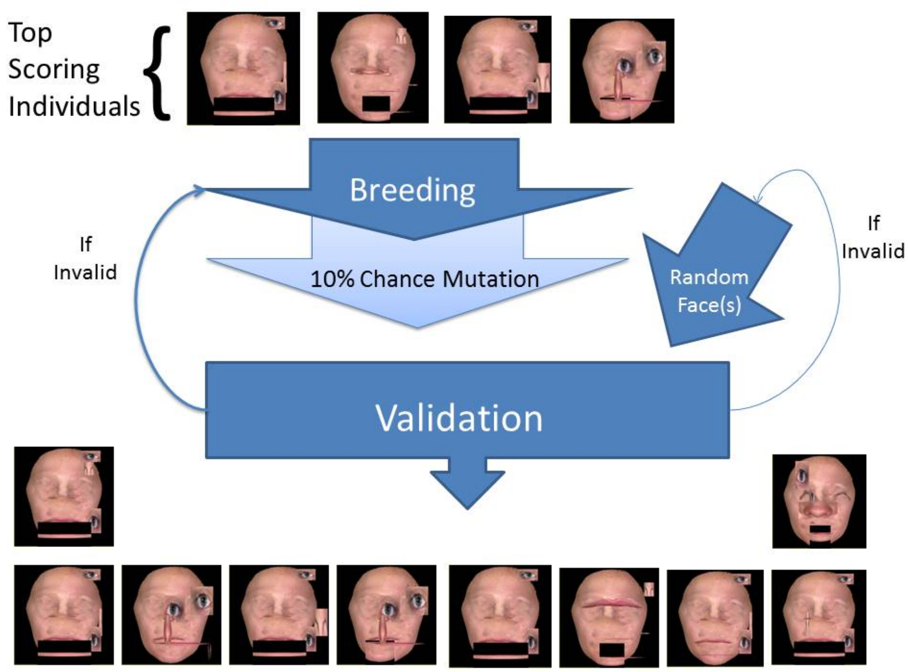
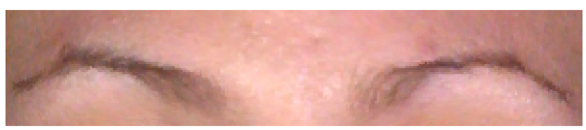


Figure 3.1: An overview of the genetic algorithm process. Top scoring individuals breed, with chance of child mutation to produce the next generation of individuals. A predetermined amount of random individuals are introduced into the population to help maintain genetic variety.

3.3.1 Chromosomes

The genes in this genetic algorithm encode for different portions of a 2D image of a face, as described in the design chapter. The partitioning of the face is inspired by the sections shown in [19]. Each section is removed from the face and placed in its own bitmap file for separate manipulation. Each portion of the face is modeled in an object-orientation manner, with each portion claiming a java class of the same name. These include eyebrows, left eye, right eye, nose, lower jaw, lower lip, and upper lip. Each class derives from an abstract base class called FacialFeature, and a face is an array of said FacialFeature objects. Each facial feature object has an id, an x coordinate, an y coordinate, the original dimensions of the feature in each y and x directions, and a scale factor for each direction to be divided by 100 at the time of transformation of the image. Some FacialFeature objects, such as the Nose and Eyebrow objects, also have the option to chose between two bitmaps. These bitmaps differ in the fact that one is masks out skin not essential to the feature to give a more realistic look when moved around the face, while the other maintains a blocky appearance. An example of this is shown in Figure 3.2.



(a) Block-like section for eyebrows

(b) Eyebrows masked out in blue.

Figure 3.2: Eyebrows in two different styles. The blue areas in the masked version is not rendered.

Although internally the values per FacialFeature are stored as member data integers, for the purposes of the genetic algorithm, the FacialFeature objects are stored in 42 bit strings.

Each encoding of a FacialFeature object begins with a four bit id, as seen in Table 3.1. The Nose object has the ID five. After that there comes two ten bit strings. The first

Table 3.1: Facial Feature Encoding - Example: Nose

ID	X	Y	Scale X	Scale Y	Image Selector
0101	0011000100	0100000010	01100100	01100100	01

encodes for the x position and the second encodes for the y position. The next two eight bit strings are the scale factors in the x and y directions. Here we show a scale factor of 100 percent, meaning that the images remain their original size. The last two digits indicate which image to use for the object. Those objects without more than one image to choose from have two zeros as their last digits.

Although the system runs smoothly, it is not without its limitations. Although error checking has been implemented, certain edge-cases still appear when producing valid media. Also, due to using a bitmap, a lot of the features needed to be relative to this specific image, and would not be applicable to other images. The transformations done on the image are also limited by the bitmap format, lacking control of elements relative to the base face.

The masking technique used in some of the FacialFeature images fell short in other FacialFeature objects which images contained a larger variety of colors. Due to some changes that occur while saving the image, no matter how pure a primary color used to mask the image, there would be some accidental blending along the edges of the unmasked portion of the image. This resulted in a very noticeable outline. To avoid this, a comparison between primary colors was made instead of a direct value check. For instance, blue was used to mask out unwanted areas. When processing the image, the blue content of the pixel would be compared to the red content, since the blue area would be more blue than red. This worked fine with areas containing high red values. However an issue was faced when trying to process more colorful images. This was seen the most with the eye objects. Since the eyes are blue and white, and the skin mostly of redder hues, masking for any of the primary colors and then comparing it to another would result in accidental omission of part of the image.

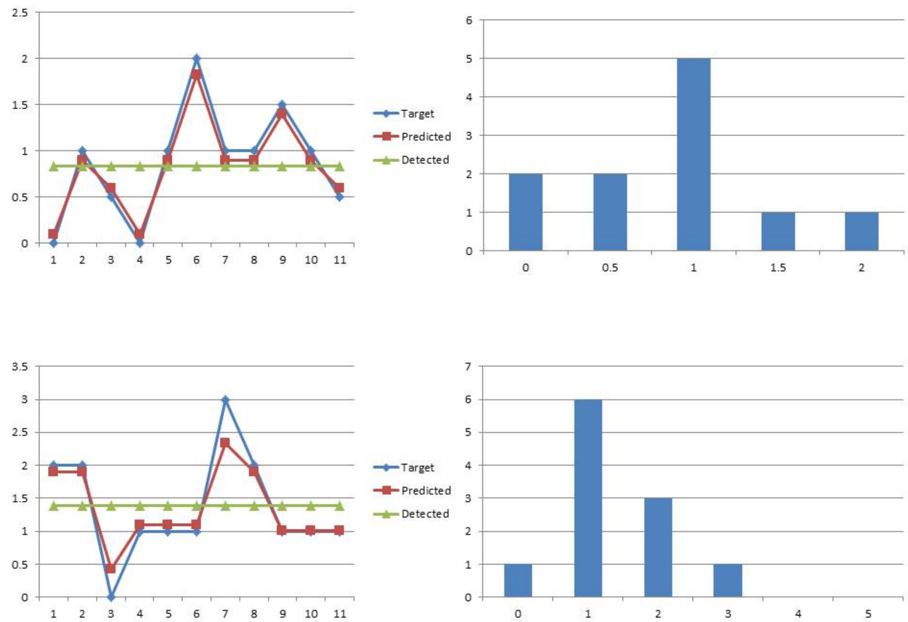
Chapter 4

Analysis

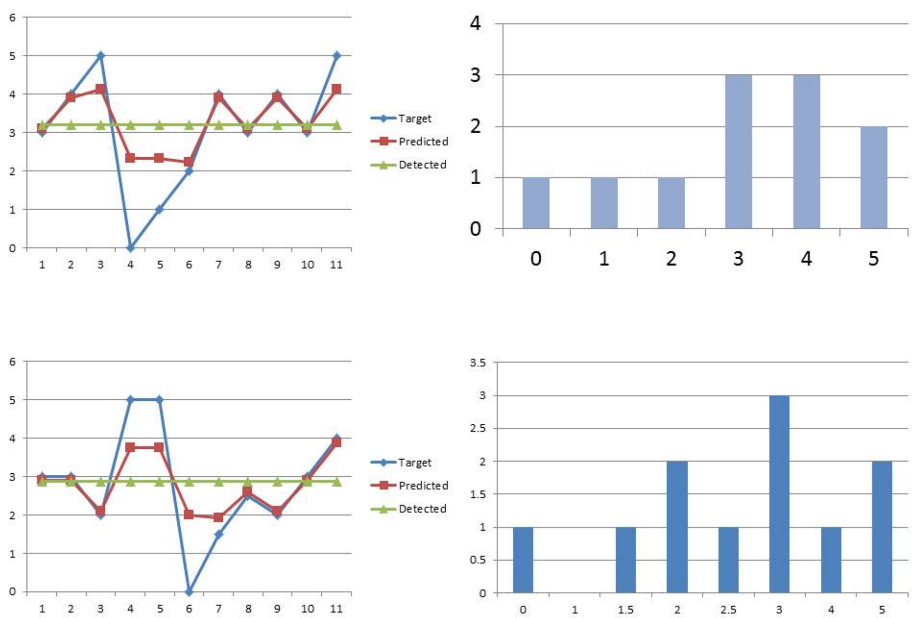
The problem this thesis addresses is complex in the fact that it requires favorable results for multiple hypotheses before the main hypothesis can be addressed. The central hypothesis to this thesis is that fear-eliciting media produced procedurally will be unique to the user. Before one can answer this question, one must answer the following: Can fear be detected using a very small sample size, and is it possible for images to become scarier using the genetic algorithm method.

4.1 Fear Detection

The first issue, of fear detection, was validated using a leave-one-out method on the data gathered in the first stage of the experiment, artificially duplicated due to the low sample size. The expectation was not an exact match, but rather that each sample would be rated correctly relative to each other, for the precise amount of fear induced is not required to discern whether one item is scarier than another. Figure 4.1 shows the results of this analysis alongside the distribution of the samples. The blue line indicates the scores given by the subject after viewing each of the film clips. The red line is the predicted score based on the physiological data gathered from the subject wile viewing each of the film clips. Scores closer to the mean found better matches in the regression than those on either end of the spectrum, which output value seemed to be pushed towards the mean in the regression. Overall, each regression was a reasonable fit for relative fear intensity detection.



(a) Data from subjects 7 (below) and 11 (above). The detected signals while viewing the face are low.



(b) Data from subjects 5 (above) and 9 (below). The detected signals while viewing the face are high.

Figure 4.1: The graphing of the user provided target scores against the scores computed and the detected scores while viewing generated material. The distribution of reported scores is shown on the right.

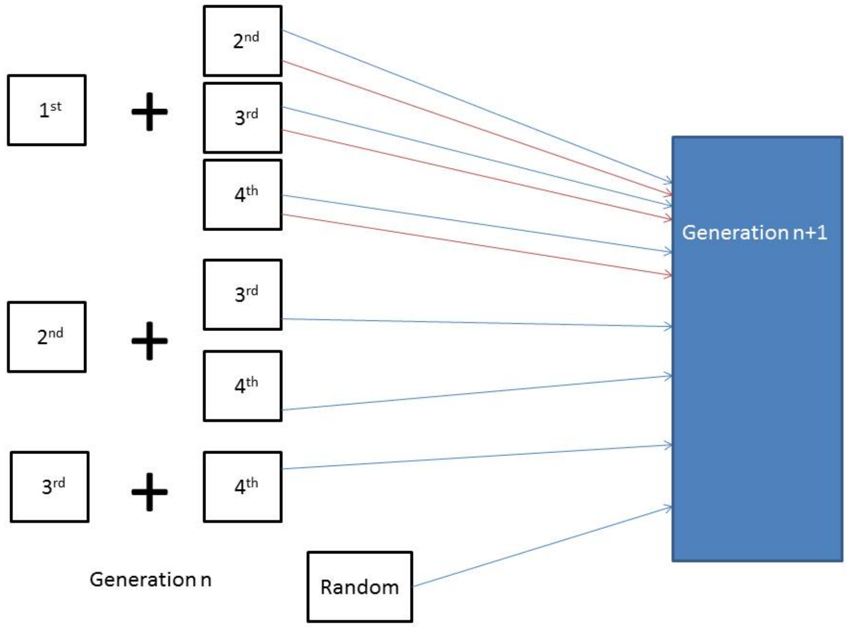
However, when shown the procedurally generated images, there was no significant difference between faces. Every subject's physiological data would indicate the faces to be a low scoring item, and each face would be exactly the same, as shown by the green line in Figure 4.1a. It should be noted that the faces did not score the subject's regression's equivalent of a zero, but rather around 1 for two subjects and just above 3 for the other two. Those that scored much higher had higher averages for their training data than those who's data returned around a one. It is to be noted that the subjects that scored higher on average identified themselves to have emotional disorders that could contribute to higher scoring. This data is displayed in 4.1b. The scores of the procedurally generated images are reflected by the horizontal line. This results could be caused by not having a robust enough generative process to produce strong and varied responses, or that noise masked the subtle fluctuations. Given that, validation of the second question needed a different method.

4.2 Increase in Fear-Elicitation of Generative Media

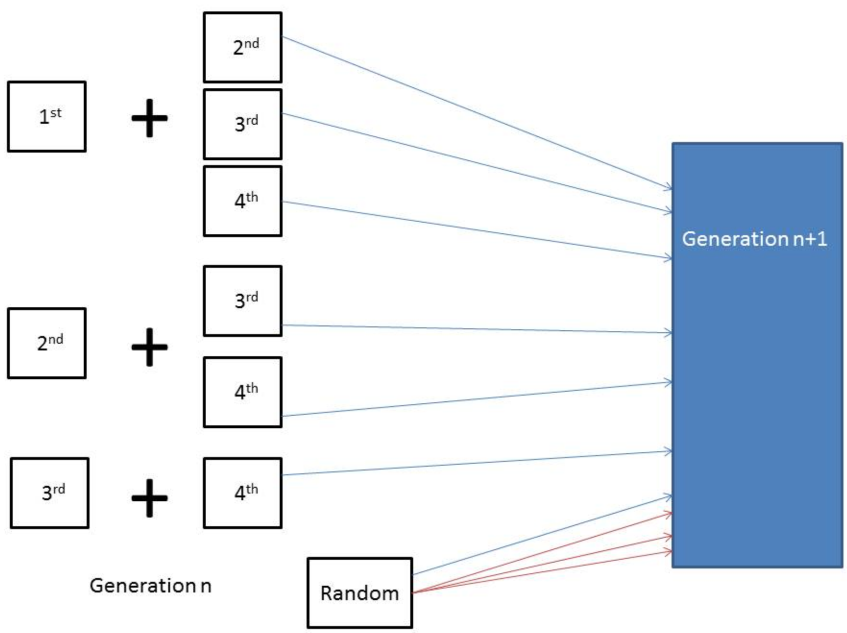
The second half of the subjects were asked to rate the faces just as they had the movie clips in stage one. This was done with the thought that the subject's direct answers would be more reactive to subtle differences in the faces than the generated score form the physiological data. The subjects did in fact give scores that differed from face to face, and in the end, there was a noticeable trend towards higher scoring faces by the end of the fifth generation. Three different variations of the genetic algorithm were used. Two of the major differences between the methods are shown in Figure 4.2.

4.2.1 Variation One - Top Four Unique Scores Chosen, One Random Individual Introduced

The first variation used was an implementation that picked the top four uniquely scoring faces to pass on genes, with one random face in every generation after the first. This caused genetic variability that may not be seen in other forms of the algorithm. The scores from the two trials (ID 6 and ID 12) that used this implementation showed the greatest percentage



(a) The method of propagation for variations that use only one random individual (variation one and two). The fittest individual breeds twice with each of the other top four individuals.



(b) The method of propagation for variation three that use four random individuals. The fittest individual breeds once with each of the other top four individuals.

Figure 4.2: Propagation methods.

of increase overall, with increases of 82.24 percent and 41.54 percent respectively. More testing will be needed to validate that the implementation is the cause of this increase. Subject 12's scores were updated in accordance with identified scoring methodology. Subject 12 would mark repeated faces as a one when the score should be the same for repeated faces. Table 4.1 shows the scores for the first and the fifth generations. Those asterisked are scores that were updated. Figure 4.3 shows the change of the average score through the five generations. A T-test was run on the change of the mean from the first to the fifth generation. ID 6's p-value ended up being 0.000, with 13 degrees of freedom. ID 12's p-value ended up being 0.028, with 13 degrees of freedom. It can be said with at least 97 percent confidence that the increase of scores between the two generations is statistically significant.

Table 4.1: Scores for the First and Fifth Generations for Subjects 6 and 12

ID 6 Gen 1	ID 6 Gen 2	ID 12 Gen 1	ID 12 Gen 5
1.2	2.0	1.0	5
1.5	1.9	1.0	5*
0.1	2.0	3.0	5*
1.3	2.1	5.0	5
1.2	2.1	5.0	4
1.0	2.3	2.5	5*
1.1	1.9	1.0	5*
1.5	1.8	5.0	5*
0.8	1.5	5.0	2
1.0	1.9	4.0	5

4.2.2 Variation Two - Top Four Scores Chosen, One Random Individual Introduced

The second variation picked the top four highest scoring faces to pass on genes, with one random face in every generation after the first. This variation focused on the genes of those that scored highest only, where as the first variation allowed lower scoring images to pass on traits. The scores from the two trials (ID 3 and ID 14) that used this implementation showed increases of 27.27 percent and 25.16 percent respectively. Subject 14's scores were

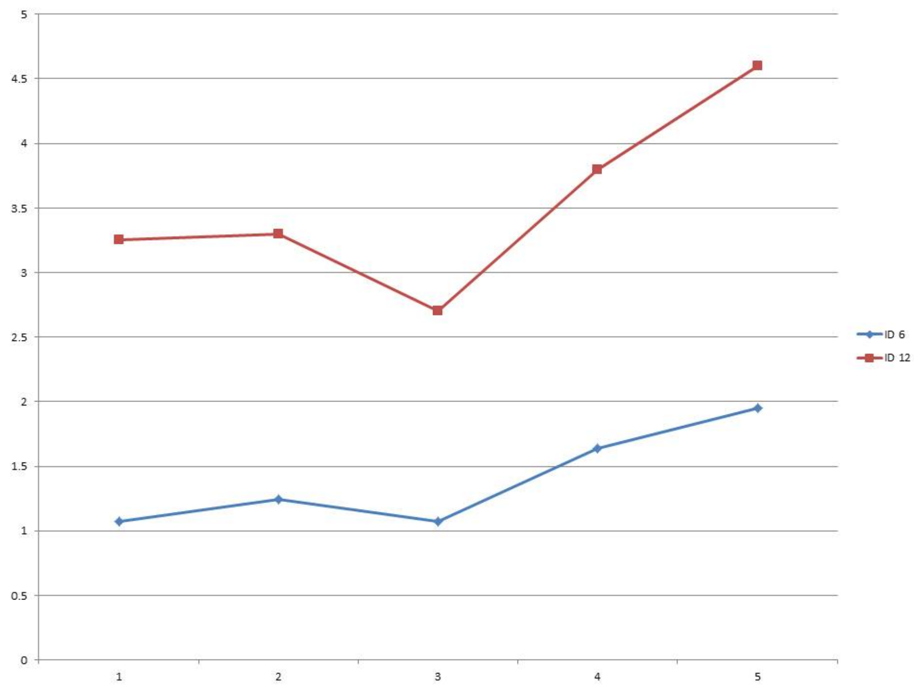


Figure 4.3: The average fear scores for each generation of ID 6 and ID 12. Note that generation 5 of ID 12 is the only adjusted data set.

updated in accordance with identified scoring methodology. Subject 3 would mark repeated faces as a one when the score should be the same for repeated faces. Table 4.2 shows the scores for the first and the fifth generations. Those asterisked are scores that were updated. Figure 4.4 shows the change of the average score through the five generations. A T-test was run on the change of the mean from the first to the fifth generation. ID 3's p-value ended up being 0.153, with 17 degrees of freedom. The null hypothesis that the means for the generations are the same for ID 3 cannot be disproven. ID 14's p-value ended up being 0.016, with 17 degrees of freedom. It can be said with at least 98 percent confidence that the increase of scores between the two generations is statistically significant.

4.2.3 Variation Three - Top Four Scores Chosen, Four Random Individual Introduced

The third variation picked the top four highest scoring faces to pass on genes, with four random face in every generation after the first. This variation had potential to provide

Table 4.2: Scores for the First and Fifth Generations for Subjects 3 and 14

ID 3 Gen 1	ID 3 Gen 2	ID 14 Gen 1	ID 14 Gen 5
1.0	2.0	1.0	2.2
1.0	2.0	1.0	2.3
0.0	1.0	1.5	2.2
1.0	1.0	1.5	2.0
1.0	1.0	1.3	1.9
2.0	2.0*	1.6	1.8
1.0	2.0*	1.4	1.6
2.0	0.0	2.0	1.5
1.0	1.0	2.0	1.5
1.0	2.0*	2.2	2.4

more new traits every generation due to the four randomly produced faces, potentially increasing genetic diversity. The scores from the two trials (ID 15 and ID 16) that used this implementation showed increases of 33.21 percent and 19.34 percent respectively. Subject 15's scores were rounded to the nearest tenth.Table 4.3 shows the scores for the first and the fifth generations. Those asterisked are scores that were updated. Figure 4.5 shows the change of the average score through the five generations. A T-test was run on the change of the mean from the first to the fifth generation. ID 15's p-value ended up being 0.004, with 15 degrees of freedom. ID 16's p-value ended up being 0.035, with 12 degrees of freedom. It can be said with at least 96 percent confidence that the increase of scores between the two generations is statistically significant.

4.2.4 Concluding Remarks for the Increase in Fear-Elicitation of Generative Media

When all of the subject's trials are taken into account, we find that there is, on average, a 107.44 percent increase from generation one to generation five. The average score given to the faces of generation one was a 1.21. The average score for the fifth generation is 2.51. A T-test was performed on the data in total and a p-value of .001 was generated with 117 degrees of freedom. This means that with 99.9 percent certainty we can reject the null hypothesis that the averages between the two generations for all subjects is the same and

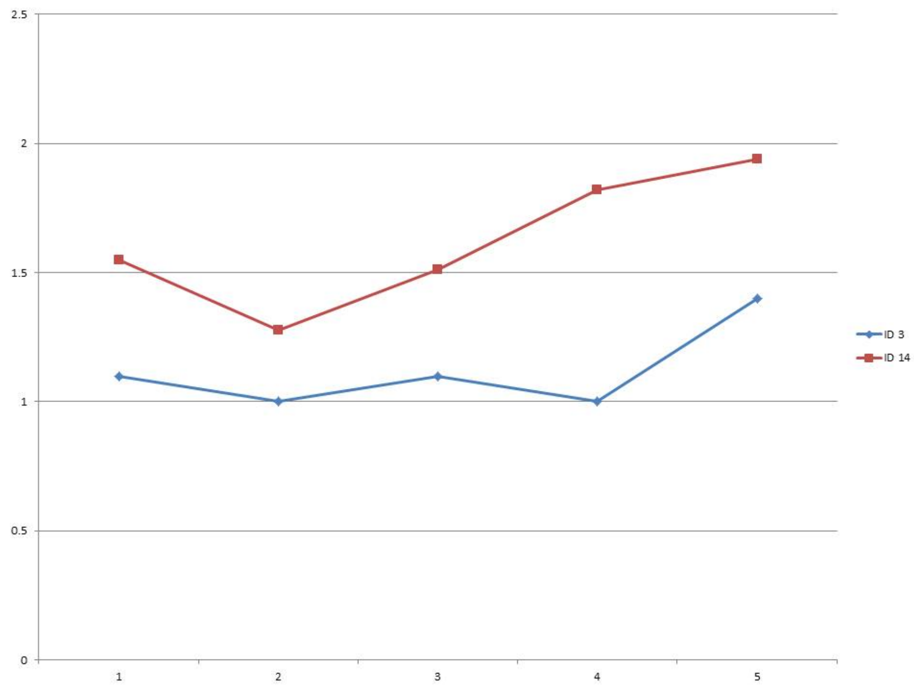


Figure 4.4: The average fear scores for each generation of ID 3 and ID 14. Note that generation 5 of ID 3 is the only adjusted data set.

accept that there is a significant increase in score from generation one to generation five. Thus our hypothesis that the generated material can become scarier is confirmed. It is to be noted, however, that many subjects found the images more unsettling or creepy versus scary. This discrepancy will be explored in the next chapter.

4.3 Uniqueness of Generated Media

With the two hypotheses showing favorable results, we can address the main hypothesis. We compared the top scoring face from each of the six subject's generation five outputs for similarity. If there is more than one face with the same score, the first face shown in the generation of that score is be used. The faces are shown in Figure 4.6. Each of the pair of faces from each of the genetic algorithm variations is compared. Each face is be masked such that facial features are blue, and non-facial feature parts of the image are tinted red. Each pair of faces is then summed together, such that the blue areas of each of the faces

Table 4.3: Scores for the First and Fifth Generations for Subjects 15 and 16

ID 15 Gen 1	ID 15 Gen 2	ID 16 Gen 1	ID 16 Gen 5
2.0	3.0*	2.5	1.9
2.0	3.3	1.2	2.1
1.0	2.7	1.2	2.2
3.0	3.4	1.7	1.7
3.0	2.9	1.8	1.8
2.0	3.8*	1.8	2.2
3.0	3.9	1.5	2.1
3.0	3.2	2.3	1.8
2.0	2.5	1.3	1.7
3.0	3.3*	1.2	2.2

appear on one face. This face is then compared to the face with the lower level of blue. The summed images are shown in Figure 4.7. The statistics for each of the faces are shown in Tables 4.4, 4.5, and 4.6. In the end, a pairwise T-test was performed between the number of blue pixels in the images with a lower number of blue pixels and the number of pixels of the summed images for each image pair. The result was a p-value of a .002 for a one-tailed test. Given this, we can reject with a 99.8 percent confidence that the null hypothesis that the low-blue image and the summed image have the same number of blue pixels and accept that the summed image has a significant amount more blue pixels. This means that there is an amount of different blue pixels in the other image of the pair, which would suggest that this image is significantly different from the other low-blue image.

Table 4.4: Number and percentage of blue pixel for the images from ID 3, ID 14 and the summed image of the two.

	ID 3	ID 14	Summed
Blue pixels	65217	60317	99681
Percentage	24.878	23.009	38.025

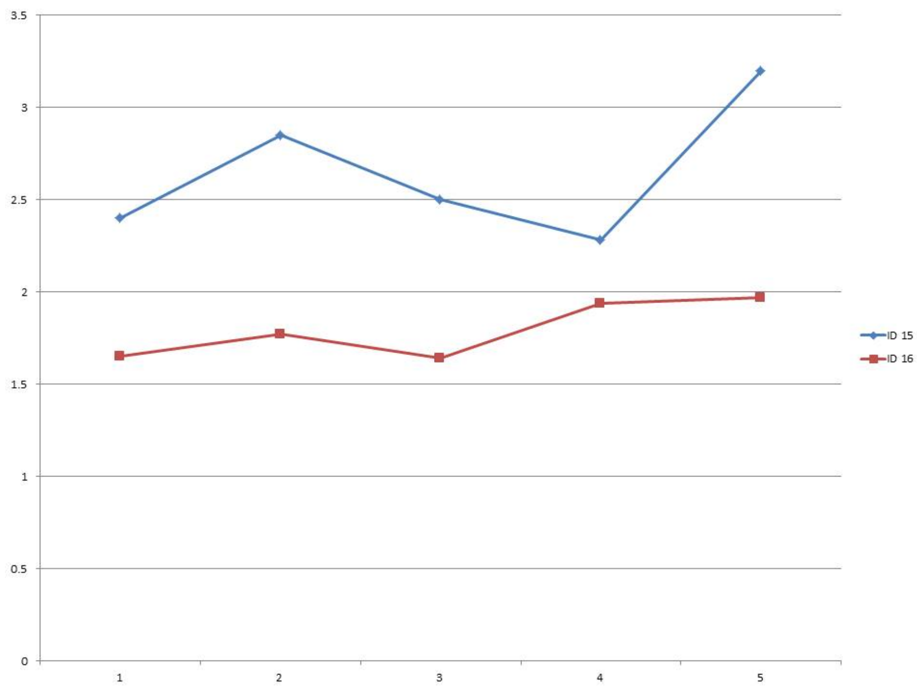


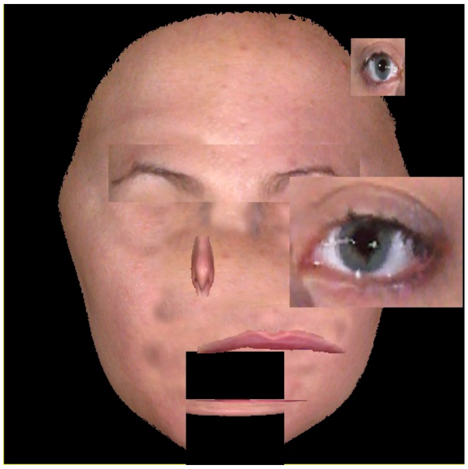
Figure 4.5: The average fear scores for each generation of ID 15 and ID 16.

Table 4.5: Number and percentage of blue pixel for the images from ID 6, ID 12 and the summed image of the two.

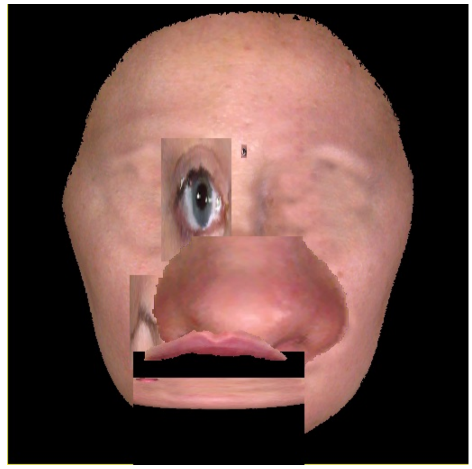
	ID 6	ID 12	Summed
Blue pixels	58713	27246	74850
Percentage	22.397	10.394	28.553

Table 4.6: Number and percentage of blue pixel for the images from ID 15, ID 16 and the summed image of the two.

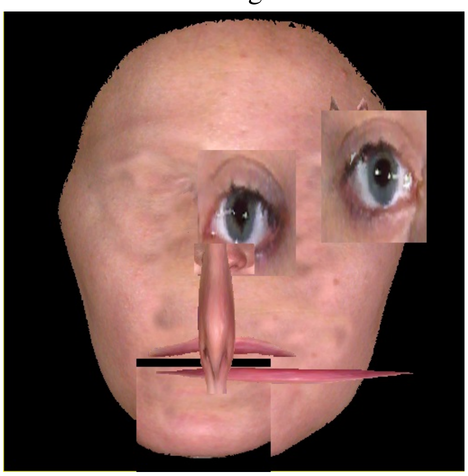
	ID 15	ID 16	Summed
Blue pixels	71422	45755	94897
Percentage	27.245	17.454	36.189



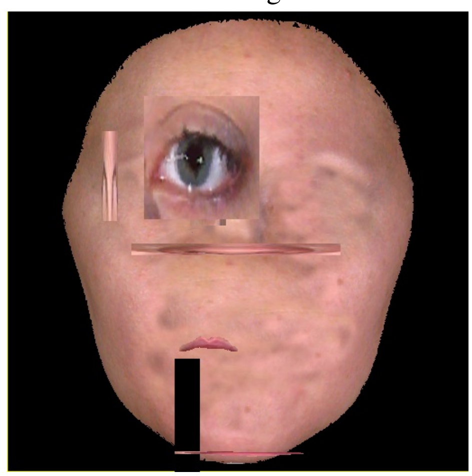
(a) The fittest face of generation 5 for ID 3 with a fear rating of 2.



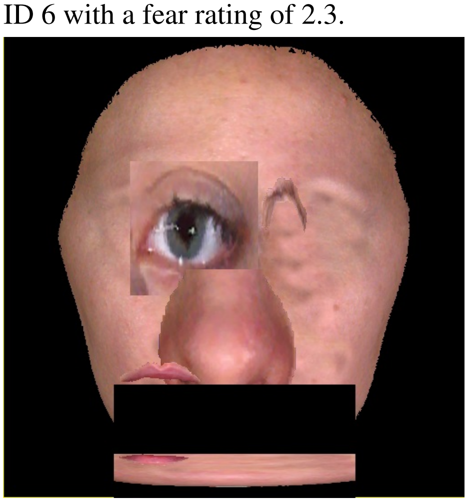
(b) The fittest face of generation 5 for ID 14 with a fear rating of 2.4.



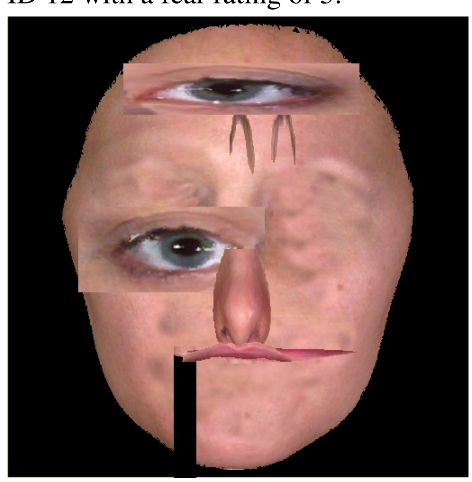
(c) The fittest face of generation 5 for ID 6 with a fear rating of 2.3.



(d) The fittest face of generation 5 for ID 12 with a fear rating of 5.

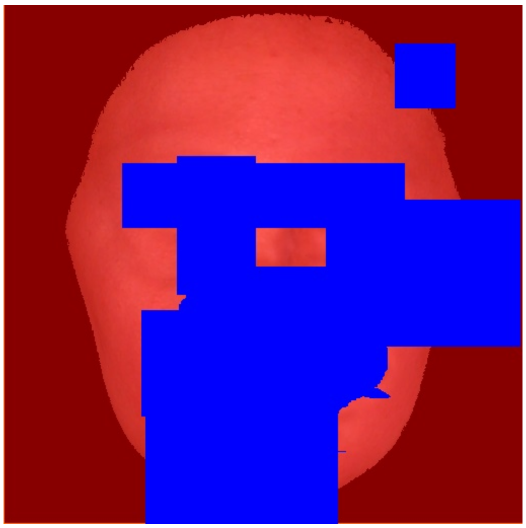


(e) The fittest face of generation 5 for ID 15 with a fear rating of 3.9.

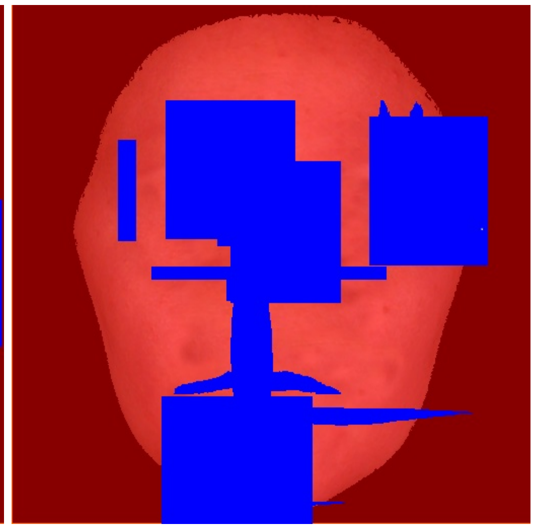


(f) The fittest face of generation 5 for ID 16 with a fear rating of 2.2.

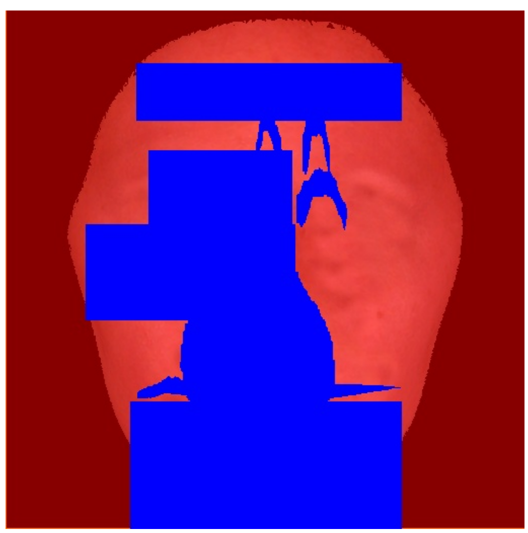
Figure 4.6: The fittest faces of generation 5 as rated by each of the subjects. Each face is to be compared with the one next to it.



(a) The overlay of the fittest faces from ID 3 and ID 14.



(b) The overlay of the fittest faces from ID 6 and ID 12.



(c) The overlay of the fittest faces from ID 15 and ID 16.

Figure 4.7: The overlay of each pair of faces used to analysis the differences between faces in the pair.

Chapter 5

Conclusions

5.1 Current Status

The system built was able to fulfill the hypothesis, with a few reservations discussed in the next section. The system is able to detect fear and create unique stimuli that gets scarier through each generation. However, we can only use the word ”scarier” here loosely, since the score used to obtain such images was from the regression trained to the physiological data gathered while subjects viewed fear-inducing material, but from user rating. Many of the subjects noted that the images shown were not scary, but rather creepy or unsettling. Although this is not exactly what was intended, it is a start. Another interesting piece of information gained from the participants of the user-rated trials was that each of the subjects found different aspects of the face disturbing, leading to the now confirmed differences in the faces. One subject cited the lack of eyes to be a large factor in rating, where as another subject indicated that larger and more abundant eyes increased their rating instead. Others still indicated that faces that were only slightly off in size or position of features increased the given score. Users also indicated that if the room were darker, context given, or faces were not repeated, a stronger response would have been triggered.

5.2 Future Work

One of the hypothesized reasons why the score generated from the physiological data while viewing the faces did not differ from face to face is that the stimuli was not intense enough relative to the training data to make much of a difference. The training data was acquired

while using movie clips, which included sound, motion and context that the face images did not. These were not in the scope of this project, but if a system that could produce animated stimuli containing audio, it would be very interesting to see what would be produced. This also addresses many of the comments obtained from the subjects saying that the faces were not fear-eliciting as much as they were creepy. The more complicated media may be able to produce the desired affect more readily.

An uncanny response can be gained from audio. The fear-eliciting comes from hearing what would be familiar sounds in relevant contexts, but rather these sounds are slightly off from what is expected in such a way that the threat cannot be discerned [21]. A common instance found in horror movies is the use of sounds that mimic animal cries, a natural alerting sound for humans. These sounds however are manipulated to hide the threat while maintaining the alerting power. By playing up fearful responses gained through evolution from threatening sounds, paired with a fear-inducing image, a rich fear experience can be obtained. Audio seems a fair bit more popular as an output from a genetic algorithm, most studies focusing on procedural music, as in [27]. For our purposes, something as complex as a fully on musical composition is above and beyond. It has been found that in such studies, such as [14] and [35], focus on qualities indicative of more complicated pieces, such as pitch, tempo, and rhythm.

Animation is another area that has seen its share of genetic algorithms. Many of them place fitness on the ability of n-pedal characters to walk [13] [28] or avoid collisions [53]. Using the given faces, it could be conceivable to give it motion or to morph between the different versions of the faces through the generations. The third factor lacking in the face images as compared to the movies is context, and that is a bit harder to produce procedurally through genetic algorithms, if at all possible. One may need to give context in a different way and use audio and animation to augment the experience.

Another good addition to this project would to allow it to be a online system, meaning that it would react to the physiological data in real time without a pause to process data. It was difficult to sync data using multiple programs to collect the data. The original intent

was to use a single program to collect the data from all the modalities, but this type of program was not available. It could have been possible to write a few more scripts to aline the data, but such was not done. It, however, still stands as an alternative to the single program.

Although it was researched and very promising, this study could not incorporate GSR data due to a malfunction with the sensors available. The fear detection may have been crisper with its inclusion. Cleaning the data for noise may have also helped in this, but it was forgone due to time constraints. It was also thought that some noise might actually be indicative to fear responses, such as jumping or tense of the body or cringing of the face. Because this thesis's intent was not to pinpoint the body's reaction to fear-stimulus, but rather to ascertain if the subject is in a state of fear overall, it was not pressing to be strict on processing the data.

Once this system is streamlined and made real-time, it could potentially be used in a variety of applications. Video games and other entertainment media are immediate thoughts. Changing the target emotion opens up more possibilities, such as for reactive product design. Potential uses in physiological therapy could also be explored, such as diagnosing and treating various mood disorders. The observation made earlier about differing data from the two participants with mood disorders leads one to believe that the research potential in this area, as well as the others previously mentioned, is promising.

Bibliography

[1] The 25 Most Disturbing Scenes in Horror - IGN.

[2] Attention Tool - Biometric Research Platform |.

[3] Nevermind | The Greatest Enemy is the One Inside Your Head.

[4] scikit-learn: machine learning in Python scikit-learn 0.16.1 documentation.

[5] Faiza Abdat, Choubeila Maaoui, and Alain Pruski. Bimodal system for emotion recognition from facial expressions and physiological signals using feature-level fusion. In *Computer Modeling and Simulation (EMS), 2011 Fifth UKSim European Symposium on*, pages 24–29. IEEE, 2011.

[6] Foteini Agrafioti, Dimitrios Hatzinakos, and Adam K Anderson. Ecg pattern analysis for emotion detection. *Affective Computing, IEEE Transactions on*, 3(1):102–115, 2012.

[7] Omar AlZoubi, Sidney K D'Mello, and Rafael A Calvo. Detecting naturalistic expressions of nonbasic affect using physiological signals. *Affective Computing, IEEE Transactions on*, 3(3):298–310, 2012.

[8] T. H. Aspiras and V. K. Asari. Log power representation of eeg spectral bands for the recognition of emotional states of mind. pages 1–5. IEEE, 2011.

[9] Jeremy N Bailenson, Emmanuel D Pontikakis, Iris B Mauss, James J Gross, Maria E Jabon, Cendri AC Hutcherson, Clifford Nass, and Oliver John. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International journal of human-computer studies*, 66(5):303–317, 2008.

[10] Steven Bergen and Brian J Ross. Automatic and interactive evolution of vector graphics images with genetic algorithms. *The Visual Computer*, 28(1):35–45, 2012.

[11] Terrence Brown, Gholamreza Fetanat, Abdollah Homaifar, Brian Tsou, and Olga Mendoza-Schrock. Human emotion detector based on genetic algorithm using lip features. In *SPIE Defense, Security, and Sensing*, pages 77040I–77040I. International Society for Optics and Photonics, 2010.

[12] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. Emotion assessment from physiological signals for adaptation of game difficulty. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(6):1052–1063, 2011.

[13] Maurizio Conventi. Physically-based animation of 3d biped characters with genetic algorithms. 2006.

[14] Arne Eigenfeldt. Corpus-based recombinant composition using a genetic algorithm. *Soft Computing*, 16(12):2049–2056, 2012.

[15] YJ Eum, EH Jang, BJ Park, SS Choi, SH Kim, and JH Sohn. Emotion recognition by ans responses evoked by negative emotion. In *Engineering and Industries (ICEI), 2011 International Conference on*, pages 1–4. IEEE, 2011.

[16] Julien Fleureau, Philippe Guillotel, and Quan Huynh-Thu. Physiological-based affect event detector for entertainment video applications. *Affective Computing, IEEE Transactions on*, 3(3):379–385, 2012.

[17] Gary Garcia-Molina, Tsvetomira Tsoneva, and Anton Nijholt. Emotional brain–computer interfaces. *International Journal of Autonomous and Adaptive Communications Systems*, 6(1):9–25, 2013.

[18] Tom Garner. Identifying habitual statistical features of eeg in response to fear-related stimuli in an audio-only computer video game. In *Proceedings of the 8th Audio Mostly Conference*, page 14. ACM, 2013.

[19] Yun Ge, Bao-cai Yin, Yan-feng Sun, and Guo-dong Jing. Expansion of 3d face sample set based on genetic algorithm. *Multimedia tools and applications*, 70(2):781–797, 2014.

[20] Khadidja Gouizi, Fethi BEREKSI Reguig, and Choubeila Maaoui. Analysis physiological signals for emotion recognition. In *Systems, Signal Processing and their Applications (WOSSPA), 2011 7th International Workshop on*, pages 147–150. IEEE, 2011.

[21] Mark Grimshaw. The audio uncanny valley: Sound, fear and the horror game. 2009.

[22] Hamza Hamdi, Paul Richard, Aymeric Suteau, and Philippe Allain. Emotion assessment for affective computing based on physiological responses. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, pages 1–8. IEEE, 2012.

[23] Jennifer A Healey. Affect detection in the real world: Recording and processing physiological signals. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

[24] Jennifer A Healey and Rosalind W Picard. Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 6(2):156–166, 2005.

[25] E-H Jang, B-J Park, S-H Kim, Y Eum, and J Sohn. Identification of the optimal emotion recognition algorithm using physiological signals. In *Engineering and Industries (ICEI), 2011 International Conference on*, pages 1–6. IEEE, 2011.

[26] S Jerritta, M Murugappan, R Nagarajan, and Khairunizam Wan. Physiological signals based human emotion recognition: a review. In *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*, pages 410–415. IEEE, 2011.

[27] Maximos A Kaliakatsos-Papakostas, Michael G Epitropakis, Andreas Floros, and Michael N Vrahatis. Controlling interactive evolution of 8-bit melodies with genetic programming. *Soft Computing*, 16(12):1997–2008, 2012.

[28] Benjamin Kenwright. Planar character animation using genetic algorithms and gpu parallel computing. *Entertainment Computing*, 5(4):285–294, 2014.

[29] A Kolakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, and Michał R Wróbel. Emotion recognition and its application in software engineering. In *Human System Interaction (HSI), 2013 The 6th International Conference on*, pages 532–539. IEEE, 2013.

[30] P Lang and Margaret M Bradley. The international affective picture system (iaps) in the study of emotion and attention. *Handbook of emotion elicitation and assessment*, 29, 2007.

[31] P Lang and Margaret M Bradley. The international affective picture system (iaps) in the study of emotion and attention. *Handbook of emotion elicitation and assessment*, 29, 2007.

[32] Aleix Martinez and Shichuan Du. A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *The Journal of Machine Learning Research*, 13(1):1589–1608, 2012.

[33] Eric J Moody, Daniel N McIntosh, Laura J Mann, and Kimberly R Weisser. More than mere mimicry? the influence of emotion on rapid facial reactions to faces. *Emotion*, 7(2):447, 2007.

[34] C Muhl and Dirk Heylen. Cross-modal elicitation of affective experience. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–12. IEEE, 2009.

[35] Noriko Okada, Mitsunori Miki, Tomoyuki Hiroyasu, and Masato Yoshimi. Classified-chime sound generation support system using an interactive genetic algorithm. In *Artifical Intelligence and Soft Computing*, pages 173–180. Springer, 2010.

[36] Emily L Parks, So-Yeon Kim, and Joseph B Hopfinger. The persistence of distraction: A study of attentional biases by fear, faces, and context. *Psychonomic bulletin & review*, 21(6):1501–1508, 2014.

[37] Panagiotis C Petrantonakis and Leontios J Hadjileontiadis. Eeg-based emotion recognition using hybrid filtering and higher order crossings. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

[38] Panagiotis C Petrantonakis and Leontios J Hadjileontiadis. A novel emotion elicitation index using frontal brain asymmetry for enhanced eeg-based emotion recognition. *Information Technology in Biomedicine, IEEE Transactions on*, 15(5):737–746, 2011.

[39] Gregory J Quirk. Fear. *Neuroscience in the 21st Century: From Basic to Clinical*, pages 2009–2026, 2013.

[40] Pierre Rainville, Antoine Bechara, Nasir Naqvi, and Antonio R Damasio. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International journal of psychophysiology*, 61(1):5–18, 2006.

[41] Kristina Schaaff and Tanja Schultz. Towards emotion recognition from electroencephalographic signals. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

[42] Suleman Shahid, Emiel Krahmer, Marc Swerts, Willem A Melder, and Mark A Neerincx. You make me happy: Using an adaptive affective interface to investigate the effect of social presence on positive emotion induction. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

[43] T. Sharma, S. Bhardwaj, and H.B. Maringanti. Emotion estimation using physiological signals. In *TENCON 2008 - 2008 IEEE Region 10 Conference*, pages 1–5, Nov 2008.

[44] T. Sharma, Tanu Sharma, S. Bhardwaj, Shweta Bhardwaj, Hima B. Maringanti, and H. B. Maringanti. Emotion estimation using physiological signals. pages 1–5. IEEE, 2008.

[45] Mariusz Szwoch. Feedb: a multimodal database of facial expressions and emotions. In *Human System Interaction (HSI), 2013 The 6th International Conference on*, pages 524–531. IEEE, 2013.

[46] Dzmitry Tsetserukou, Alena Neviarouskaya, Helmut Prendinger, Naoki Kawakami, and Susumu Tachi. Affective haptics in emotional communication. In *ACII*, pages 1–6, 2009.

[47] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. The social signal interpretation (ssi) framework: multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 831–834. ACM, 2013.

[48] Steffen Walter, Jonghwa Kim, David Hrabal, Stephen Clive Crawcour, Henrik Kessler, and Harald C Traue. Transsituational individual-specific biopsychological classification of emotions. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, 43(4):988–995, 2013.

[49] Wan-Hui Wen, Guang-Yuan Liu, Nan-Pu Cheng, Jie Wei, Peng-Chao Shangguan, and Wen-Jin Huang. Emotion recognition based on multi-variant correlation of physiological signals.

[50] Joyce Westerink, Martin Ouwerkerk, G-J de Vries, Stijn de Waele, Jack van den Eerenbeemd, and Marco van Boven. Emotion measurement platform for daily life situations. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

[51] Georgios N Yannakakis. Game adaptivity impact on affective physical interaction. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

[52] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J Rosato. A 3d facial expression database for facial behavior research. In *Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on*, pages 211–216. IEEE, 2006.

[53] Min Zou. Real-time interactive character animation by parallelization of genetic algorithms. In *Motion in Games*, pages 254–265. Springer, 2011.

Appendix A

UML Diagrams

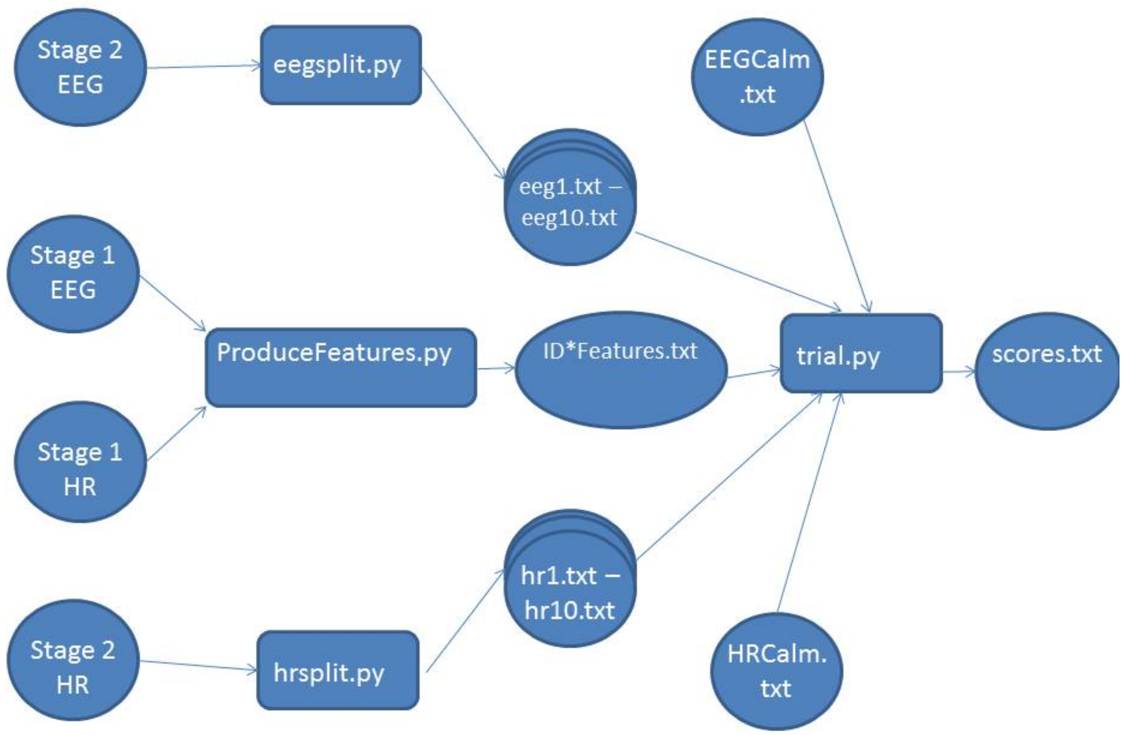


Figure A.1: This diagram shows the flow of data from the physiological signals to the end fear rating, found in scores.txt.

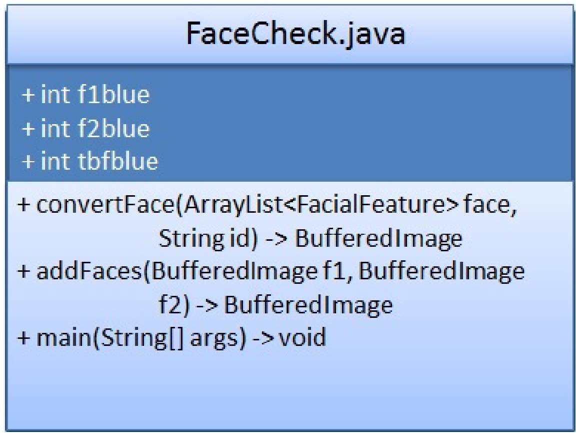


Figure A.2: The class used to check the differences between two faces.

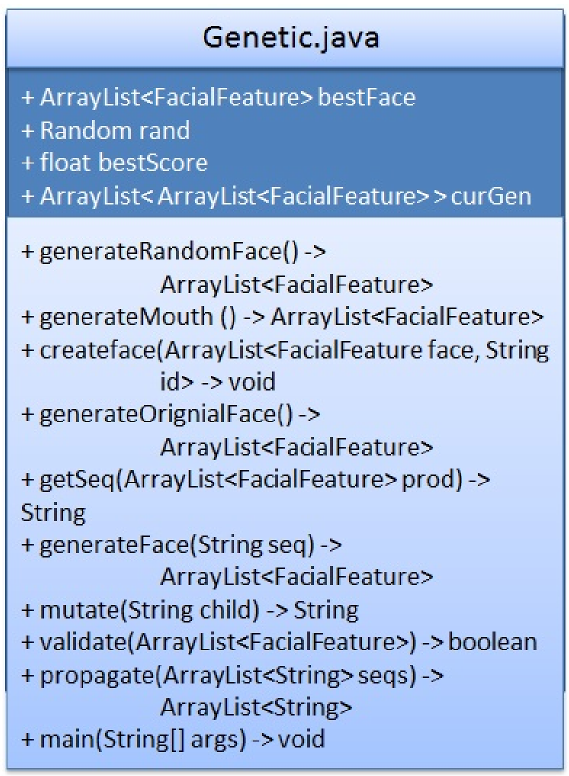


Figure A.3: The genetic class is the main driver of the media generation.

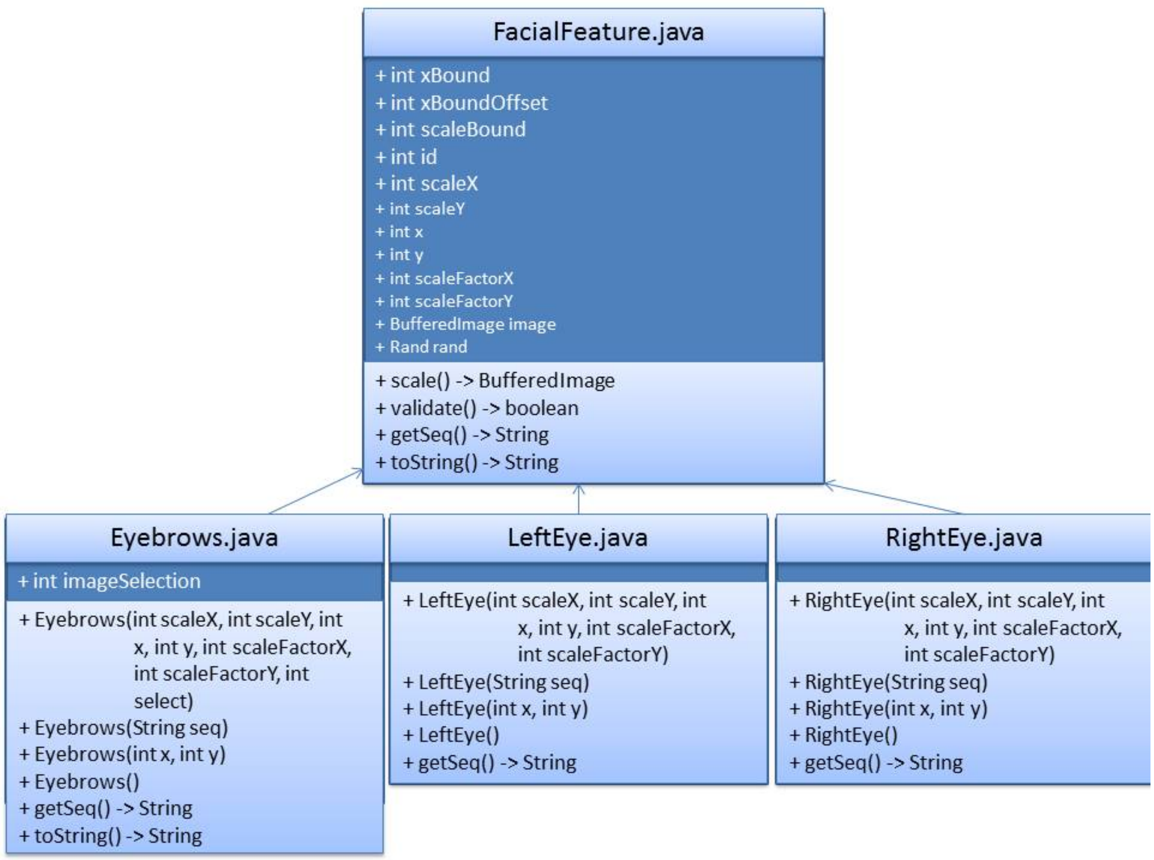


Figure A.4: FacialFeature is the base class to Eyebrows, LeftEye and RightEye.

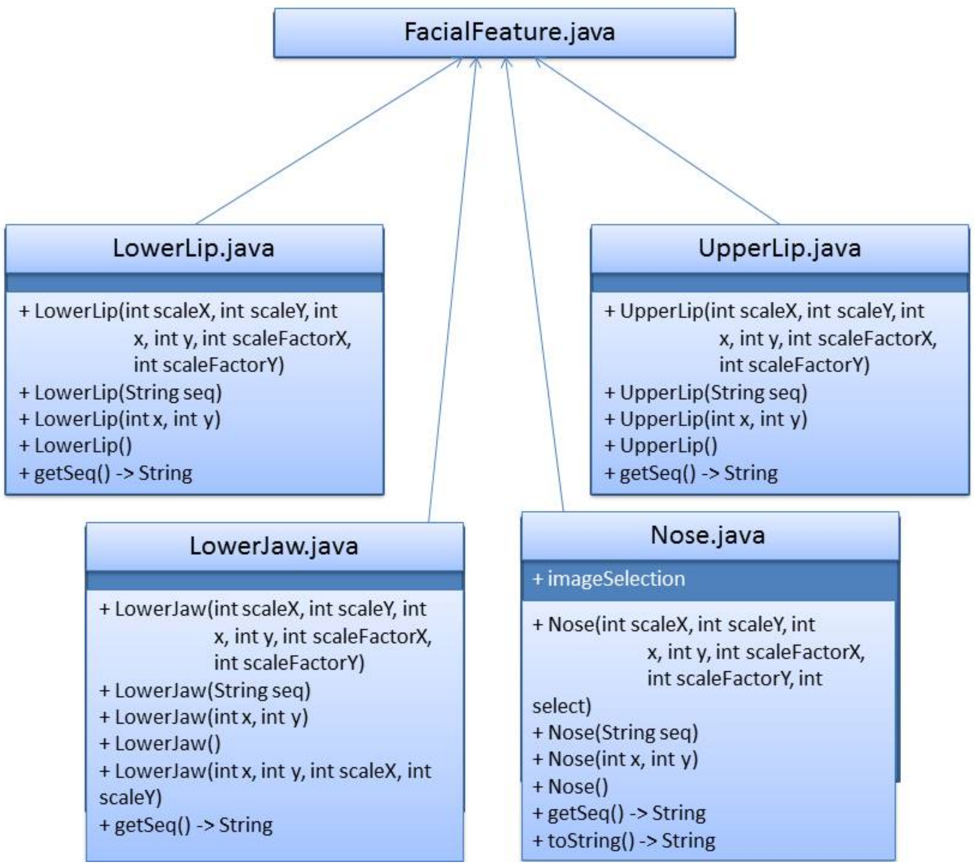


Figure A.5: FacialFeature is also the base class to LowerJaw, LowerLip, UpperLip and Nose.

Appendix B

User Manual

The following is to run the genetic algorithm to produce faces.

The program can be run in two ways, either with no command line arguments, or one.

java Genetic

will cause the program to come up with its own initial population consisting of 9 randomly generated faces and the original face.

If one runs the program with one argument such as

java Genetic base.txt

it will use the faces specified in the file. Each line depicts a single face. For more information on how each face is formatted, please refer to the toString() function found in each of the classes deriving from FacialFeature.

One the program is running, it will initialize the first population. It will wait for any input besides '0' to produce the images described by the first generation. If '0' is given at any point, the faces listed in the current generation are outputted to "out.txt". If the '0' comes before any other input in the generation, the program also quits.

After input besides '0' is received, faces are outputted to numbered faces files, such as "face1.txt". It will then wait for the input '2'. At this point, a score file named "score.txt" should be available, for it is needed to check the fitness of each of the faces. This file is formatted that each face is given a line, with the float score following the face's id. The program then waits for any other input or a '0' to repeat the process.