Group 503 presents

Marcel Schmidt
Rune Stenstrøm
Jan Olesen
Lars Hagen
Alexander Birke
David Teglbjærg

Affective Measurements in a Horror Game

Aalborg Univesity Copenhagen- Medialogy fall 2011

AudioVisual experiments - computer graphics and animation

Supervisor : Rama Hoetzlein - Co-supervisor : Stefania Serafin

Aalborg University Copenhagen

Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, Denmark

Secretary: Judi Stærk Poulsen
Phone: (+45)99402468
mks@imi.aau.dk

Semester: 5th

Title: Affective Measurements in a Horror Game

Project period:
1. september – 21. december 2011

Semester theme:
Audiovisual Experiments – Computer Graphics and Animation

Supervisor:

- Rama Hoetzlein

Co-supervisor:

- Stefania Serafin

Project group no.: 503

Members:

- Marcel Schmidt
- Rune Stenstrøm
- Jan Olesen
- Lars Hagen
- Alexander Birke
- David Teglbjærg

Synopsis:

The initial goal of this project was to create a horror game that would self-adjust to the player's emotion.

The article contains investigations on how to measure emotions through physiological response as well as through people self-reporting their emotions.

The final focus of the article is the correlation between the magnitude of self-reported fear and magnitude of physiological response of people playing horror games.

The study includes a pilot test where the galvanic skin response and heart rate of the participants were measured, while they played the popular survival horror game Amnesia – The Dark Descent. The test concluded that the overall tension of the game was too high and therefore not suited for our test.

A prototype was therefore built with focus of creating a higher contrast between calm and scary moments.

22 test participants tested the game while their physiological data was recorded. The analysis of the data concluded that scary events trigger an increase in galvanic skin conductance. However as physiological response is highly individual and due to a biased test procedure further research is needed in order to investigate this relationship.

Affective Measurements in a Horror Game

Rune Stenstroem, Alexander Birke, Jan Olesen,
Marcel Schmidt, Lars Hagen, David Teglbjaerg

December 20, 2011

Abstract

This article contains investigations on how to measure emotions through physiological response and self-reporting. The final focus of the article is the correlation between the magnitude of self-reported fear and magnitude of physiological response to scary events in a horror game. The study includes a pilot test where the galvanic skin response and heart rate of the participants were measured, while they played the popular survival horror game Amnesia - The Dark Descent. The test concluded that the overall tension of the game was too high and therefore not suited for our test. A prototype was therefore built with focus of creating a higher contrast between calm and scary events. 22 test participants tested the game while their physiological data was recorded. The analysis of the data concluded that scary events trigger an increase in galvanic skin conductance. However as physiological response is highly individual and due to a biased test procedure further research is needed in order to investigate this relationship.

Keywords:

Affective Gaming, Survival Horror, Fear Measurement

1 Introduction

I'm standing in a torture chamber. One can only suspect what has happened in this room based on the blood on the walls and the dead bodies lying on the floor. I'm tense and I can feel my blood pumping as I approach the door at the far end. Suddenly one of the half rotten corpses grabs my leg and pulls me to the floor. I scream as I press the pause key and instantly exit the game.

Above scenario could easily be a situation in a horror game where the player is simply too afraid to keep playing even though the player might enjoy other elements of the game such as the story or the puzzles. On the other hand, there also exist players who might not find the horror game scary enough and therefore lose interest.
Most of the current games use heuristic techniques to balance the difficulty, such as the player's position in a racing game or time taken to complete a task [37]. However no commercial games try to adjust itself to the emotions of the player.
The term affective gaming refers to games balanced by an artificial intelligence assessing the emotions of the player. By doing so it can continuously adjust story, visuals, sound, etc accordingly and thereby serve the player a more enjoyable experience.
We believe that affective gaming can be used to adjust how scary a horror game is so it suits the individual player. This study will therefore investigate affective gaming in relation to horror games.

Since affective games typically use physiological data to assess the emotions of the player this area will be investigated first.

2 Utilizing Physiological Response of the Player

The fact that video games has a physiological effect on players is well-known. Much research has been done in this area. For example, Segal and Dietz compared players' heart rate, blood pressure and oxygen consumption, over a 30-minute duration playing Ms. Pac-Man under laboratory conditions [35]. They concluded that when people played Ms. Pac-Man these three physiological properties increased significantly. Another example is Griffiths and Dancaster who investigated video game addictiveness. Their experiment also showed that the arousal level of participants increased during play. Griffiths and Dancaster furthermore concluded that the base level of arousal differs from one type of persons to another[10]. Research by Barlett et. al. showed that the amount of blood in a game affects the player's aggression, hostility and arousal [3].

The act of utilizing the physiological response from people is referred to as biofeedback and is often used in affective gaming. The next sections will investigate current biofeedback game research.

2.1 Implicit and Explicit Biofeedback

Kuikkanemi et. al. investigated the mapping of physiological data to in-game variables in a first person shooter game. They categorized the mapping of the physiological data into two groups. Explicit and implicit biofeedback.[21]. Explicit biofeedback refers to mapping of physiological data to parameters in the game, where the player is aware of the mapping. This could be an implementation where the player can control the run speed of the avatar by alternating his respiration rate. Implicit biofeedback, on the other hand, refers to a situation where the player is only subconsciously aware of the physiological control method and the application is controlled indirectly according the affective signals. The player is, however, still aware that his or her psychological state is being sensed. This could be alternating the run speed according to the players' heart rate.

Kuikkanemiet. al. showed that explicit feedback was more appreciated by the participants and easier to implement in regards to game play design. Implicit biofeedback is best for controlling background variables, such as environmental conditions. Their study thus showed that in order for biofeedback gaming to be effective it should not affect the player's ability to interact with the game world, but rather change variables in the environment to adjust the game.

2.2 Direct and Indirect Biofeedback

Nacke et. al. developed a 2D platform shooter game based on the findings by Kuikkanemi. A wide selection of physiological sensors was classified as either direct or indirect and mapped to different variables in the game world. Direct feedback refers to physiological parameters that a person has direct control over, such as respiration rate or muscle tone. Indirect feedback refers to physiological parameters that the player does not have direct control over, such as skin conductance and heart rate. Based on their findings Nacke et. al. argue that direct physiological biofeedback should be mapped intuitively to reflect an action in the virtual world, whereas indirect physiological input is best used to alter background variables. They also argued that physiological controllers should augment traditional controllers, rather than substitute them.[29]

2.3 Biofeed the Zombies

The most similar related project is a design study conducted by Dekker and Champion. They developed a first person shooter horror game where they altered gameplay, sound, and visuals based on the player's galvanic skin response and heart rate [4].

The two most interesting aspects of their study is that they investigate how physiological data of the player can be mapped to game elements. Secondly, the device used in their research is the same device available for our research. They conclude that the device can appropriately measure the player's level of arousal. However they choose to place the device on the player's fingers which made it frustrating to control the game by mouse and keyboard. This was the root of several error sources in their project.

There are additional problems regarding this research. Firstly they do not explain how they determine the different threshold values for when to trigger elements based on the biofeedback data. Secondly the argumentation for the specific mapping of the biometric measures to elements in the game seems rather poor. Furthermore they map indirect physiological parameters, such as heart rate or galvanic skin response, to game controls which are usually controlled directly by player input. As the research by Kuikkanemi et. al. and Nacke et. al. show this can make the player frustrated.
The test method of Dekker and Champion also seems to be biased. They tested if the biofeedback enhanced version of their game was more preferred than the normal game. The majority of test participants reported that they preferred the enhanced version. However alone the enhanced version had extra visual effects and it is therefore likely that the test participants would prefer this solely due to the enhanced visuals.

2.4 Affective Gaming

Although the study by Dekker and Champion is relevant, Gilleade et. al. argues that it is not enough to merely replace or augment the traditional control input with physiological input. In biofeedback games the player explicitly participates in controlling their physiological responses in order to control the game world. Affective gaming instead uses the physiological data to assess the player's emotional state and thereby adjusting the game to suit the player's emotions.

Computer games have always effected the emotions of players. Everyone that plays computer games knows the arising feeling of frustration or anger when losing a game or the feeling of joy when completing a difficult level [13]. However for a game to be affective it needs to propagate affective feedback. In other words, the computer needs to be an active intelligent participant in the biofeedback loop and interpret the emotions of the player based on the physiological data and react on it in an intelligent way [9].

An example of a game assessing the emotions of the player is the game called "Feed the Fish" implemented by Obaidet. al. They used facial recognition software to read the player's emotion and adjusted the difficulty of the game based on the video capture. They tested the difference in enjoyment of players playing with and without the affective system enabled. They experienced a small increase of enjoyment, but they did not get any significant results. One reason could be, as they conclude, the light conditions interfered causing lowering in the accuracy of the facial recognition software [28]. Although, this game to some extend uses the affective state of the player, it does so by utilizing the facial expression of the player. The accuracy of facial recognition software depends highly on environmental variables, such as lighting. Since biofeedback seems to be a more solid way of assessing the player's emotions, there is need for research focusing on how the physiological response of players relate to their emotions.

3 Emotions

A big conflict in emotion research is whether emotions can be categorized into discrete categories, a viewpoint held by well known emotion researcher Paul Ekman or whether they exist as points on a set of dimensions in an emotional space. This is the perspective used in almost all research in affective computing. This and the section will therefore also give short introduction of both viewpoints.

3.1 Modalities of Emotion

What is the role of emotions? There are many different opinions on this matter but the general consensus in affect research is that emotions evolved as a way to prepare the individual for upcoming events and are present across multiple modalities. This point of view is shared by Ekman. In his categorical definition of what he calls basic emotions he lists that emotions create an appropriate physiological response to the feeling. At the same time they also create appropriate facial expressions and body language to inform other what the individual is feeling [8]. An overview of Ekman's definition of basic emotions are presented in appendix A.

The general consensus is that each distinctive emotion has a unique signature across the different modalities and can be identified through these signals. Some of the signatures for fear are for example to cry, raised eyebrows and pulled together narrowed eyes, and a closed and downward looking posture[20]. These signatures can, according to Hudlicka, all be categorized under the behavioural/expressive modality. Furthermore, fear typically causes heightened galvanic skin conductance levels and raised heart rate, which can be categorized in the somatic/physiological modality [20]. Other modalities include cognitive/interpretive modality and the experimental/subjective modality [13]. Different emotions can have almost undistinguishable signatures in one modality and different signatures in another modality. This means that cross-modal measurement and comparison is a necessity if one is to pinpoint a specific emotion.

There are several ways of measuring the signatures of emotions on the different modalities. One apparent method is through self-reporting, meaning that the participant reports the emotion that they are feeling. This technique is far from problem free. First of all, according to Mandryk and Atkins, subjects are bad at self-reporting. Moreover this technique only generates data when a question is asked leaving this measuring method far from real time. Lastly self-reporting is often too disruptive and can cause lowered immersion of the player. However self-reporting is often used to validate data acquired by the other measuring methods [13, 25].

Other means of measuring include the use of video to interpret facial expressions, body language and gestures. This can give informational rich data, but in the world of computer games players tend to be, posture wise, very still. Furthermore their facial expressions tend to be very subtle and hard to detect [25]. It is only extreme game events which cause noticeable changes in facial expression and body language. There are several other non-physiological ways of measuring emotions, but the affective game research community tends to agree that the physiological measuring methods are the most suitable.

The most commonly seen test setup consists of a measure of skin conductance, covering the arousal axis and the use of EMG to measure the activity of the facial muscles to cover the valence dimension [18, 30, 12].

Due to lack of time and resources the only measuring device available for this research the IOM Wild-divine. It can measure skin conductance and heart rate respectively. These are both physiological measures and signatures found in the physiological modality of emotion. They are therefore only indicators of the level of arousal and not the person's level of valence. They alone are therefore not sufficient to measure a specific emotion of a person.

It will not be possible to implement real time data acquisition valid enough for affective computing using the IOM device. The focus of this research must therefore be modified so it suites the means of measuring we have available. In order to implement an affective horror game it also seems necessary to identify a more precise correlation between physiological arousal and fear. A review of affective studies found in an article by Kreibig et. al. shows that there has only been conducted research to test if there is a positive or negative correlation between fear and various physiological signals[20].
The goal of this study will therefore be to see if it is possible to find a more precise correlation between physiological arousal and fear. This will be done through non-real time validation by supporting the data acquired by the IOM device with self-reporting. The next section will therefore investigate different methods of how to do this.

4 Self Reporting

In the body of research regarding affective gaming and computing the most prominent way to measure affect is through the arousal-valence space and the Affect Grid scale. Since they are used in almost all research on this topic it necessary to provide an overview of these in order to explain why they cannot be used in the context of this research and why another way of measuring affect is needed.

In dimensionsal models of emotions the two most often used dimensions are valence(unpleasantness-pleasantness) and arousal(high-low) that forms two axes of what is called the Arousal Valence space of emotions. The AV space was popularized by Russell's work on the circumplex model of affect and Russell, Weiss and Mendelsohn in their article on the Affect Grid [34, 33]. Based on earlier studies Russell concluded that arousal and valence are the most important dimensions to take into account in an affect space. This led to the circumplex model of affect depicted on figure 1.

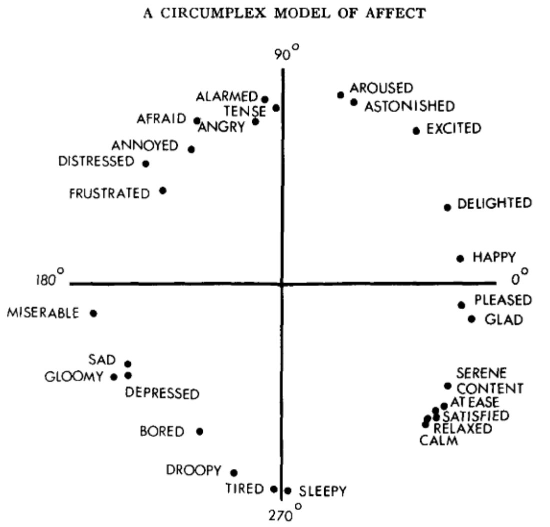


Figure 1: Russell's circumplex model of affect [33]

The Affect Grid (see fig. 2) was later developed as a way to perform self reporting of these two dimensions. It consists of a two dimensional grid with one axis for arousal and another for valence. Participants place check marks on the grid in response to different stimuli.

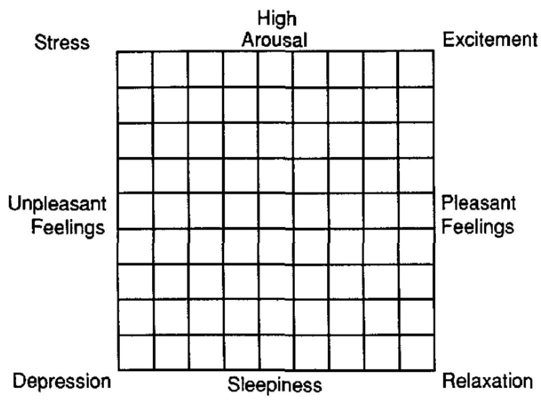


Figure 2: The Affect Grid[34]

The biggest strength of the Affect Grid is that it is very quick for a test participant to fill out. Russell, Weiss and Mendelsohn also critizise categorizing emotions since the participant can have a feeling that does not fit into the a prior categories defined by the researcher. Categorical measures based on the theories of Ekman and similar also tend to be correlated with one another. The emergent current view is that the Affect Grid is an accepted form to report emotions [19].

4.1 Criticism of the Affect Grid

Even though the Affect Grid has seen a large degree of success there are multiple complications with it. First of all are the feelings of fear and frustration the two most likely feelings to occur in the test. Fear as a result of the scary events in the game and frustration as a result of the interaction problems the IOM device is likely to cause. Both frustration and fear are likely to generate somewhat similar responses if the Affect Grid is used since both feelings would register as having high negative valence and high positive arousal levels as depicted on fig. 1. Kreibig. Et.Al. also reported that both fear and sadness maps to the same place in the AV Space indicating it is a general problem with the model[20].
Russell, Weiss, and Mendelsohn also consider arousal and valence to be orthogonal dimensions but admit that at certain circumstances they will not be. This is mostly for high levels of arousal and valence where studies have shown that a certain codependency exists [22].
Another problem with the AV space is that it can only represent one emotion at a time, other research indicates that multiple emotions can exist at the same time and even that valence can be split up into two separate dimension of the affective space [39]. The concept of having ambivalent emotions toward something is also a clear indication of this. Even the enjoyment of watching a horror film or playing a horror game supports this claim. The game or film creates a negative emotion such as fear, disgust or tension, but at the same time is enjoyable to watch.

As a result it is necessary to use another method for self reporting in the context of this study. This will be investigated in the next section.

4.2 The PANAS Scale

Instead of making a new scale to measure affect and risk it is subject to low reliability and validity another scale is sought instead. One complication with this is that it is very hard to find another scale developed to measure brief measures of affect other than the Affect Grid. Through research the only suitable scale we were able to find is the PANAS scale. The Positive Affect Negative Affect schedule abbreviated to the PANAS scale is an attempt to measure both positive and negative affect at the same time as two separate dimensions [39].
The PANAS scale consists of a series of words that describe different emotions and affective states. The participant answers how much a word describes his or her affective state by assigning it a number from 1 to 5. Each of the scales consists of a subset of these words and the mean value of these values are then the score of that affective scale.
It has been tested on large sample sizes and has proven to possess a high degree of internal consistency. It has also proven to have excellent convergent and discriminant correlations. An expanded version called PANAS-X has also been developed that additionally measures other dimensions of affect, such as fear[38]. Because of this it was chosen to use a modified version of it for the purpose of this project.

Now where a method for self reporting has been identified the iterative implementation of the biofeedback device will be explained next.

5 Implementing Biofeedback device

The implementation of the biofeedback device is done through several iterations. At the end of each iteration the current implementation is tested on a small number of persons. These tests thereby serve the purpose of refining the data processing methods so they suite the purpose of this research. Additionally the tests serve as pilot tests, since they can also pin point potential error sources. This makes it possible to find the right way to correct them before the final test. Lastly, the tests also serve the purpose of giving an initial understanding of how heart rate and skin conductance relate to the arousal of the player.

5.1 Biofeedback Device

The IOM device consists of two sensor systems. One detects galvanic skin response (GSR) by creating a small voltage across the skin of a person and measure the conductance which will vary based on the moisture of the skin.
The other system is a diode and a photosensitive resistor. The diode emits light through the finger and depending on the amount of blood in the finger, the photosensitive resistor receives different amounts of light. By measuring the resistance one can get the

heart rate variance of a person. This can then be used to calculate the heart rate.

5.2 Initial Data Processing

The skin conductance data seems to be accurate, so currently only the heart data is processed. Heart rate is typically expressed as beats per minute and hence it is measured over a period of time. There are several methods to determine the heart rate of a person. The standard way is to count all beats during the period of one minute. If the heart rate is constant this method is precise, but changes in heart rate will be smoothed out with this method and if small enough be removed. The goal of our software is to have the approximate real time heart rate of the player in beats per minute. Other means of measuring is therefore needed.

Since the heart rate data is actually data based on the amount of blood in the person's finger, a threshold is needed to check whether a peak is occurring. The current method uses the intervals between the last five peaks to estimate the beats per minute. This method is chosen as it detects fast changes in heart rate.

5.3 Test Procedure

The game chosen for the test is called Amnesia. Amnesia is a new horror game which is known as one of the scariest games ever made [27]. In each test the level of arousal of the participant is measured, while he or she is playing Amnesia. The data is hereafter analysed.

5.4 First Iteration

In the first iteration three participants played Amnesia for 30 minutes each. Meanwhile their heart rate and skin conductance are measured. There seems to be correlation between the events happening in the game and participants' level of skin conductance. The heart rate on the other hand is not accurate. It seems to vary sinusoidally. This could be due to respiratory sinus arrhythmia, which is the variation of the heart rate during respiration [40]. Another way of measuring the heart rate is therefore needed to be implemented before next iteration.

Another problem relates to the sample rate of the device. Currently, the data retrieval software uses the sample rate of the device to write the data to a log file. This means that a person playing for 30 minutes generates 60.000 data points. Collecting data at this rate is not necessary and it results in the use of too much computational power to process the data. Before the next iteration the software will be modified to collect less data.

5.5 Second Iteration

In this iteration the sample rate of the software has been changed to 1 sample per second. Additionally a new way of measuring the heart rate has been implemented. This method calculates the amount of beats during the last minute every two seconds. It thereby utilizes the standard way of calculating beat per minutes, but reports it every two seconds. In this iteration four persons participated in the test. Skin conductance still seems to correlate with what they are experiencing, so a sample rate of 1 sample per second work appropriately. Heart rate on the other hand is still not accurate, but this is due to the sensor type used by the device. If the participants are moving the fingers the device is attached to, additional light can enter the photo resistive sensor causing a peak in the signal. Some filtering method to remove these erroneous peaks in the signal needs to be implemented.

5.6 Third Iteration

In the third iteration the data retrieval software has been rewritten to forward the data via UDP to MAX/MSP. A MAX/MSP patch then filters the heart rate signal. This is done by two low pass filters in series with the purpose of smoothing the signal. A threshold value hereafter determines if a peak is occurring. Additionally the time between each peak is calculated. If the time between them is not realistic the new peak is considered an erroneous peak. Currently this threshold is set to 300ms.

This setup was tested on three test persons. The heart rate signal gives a steady reading, but there does not seem to be correlation between the events in the game and the change in heart rate. This could be random, since the sample size is only 3 persons, so a more elaborate test is needed to assess the usefulness of the heart rate signal.

5.7 Limitations of Amnesia

An overall problem regarding the initial test is that Amnesia is developed to create a constant atmosphere of suspense. This means that mapping game events to fluctuations in the signal will be hard and biased. In order to accurately compare game events to the signal there must be bigger contrast between event causing arousal and events not heightening the player's level of arousal. In order to create such a test environment it is necessary to create our own prototype.

6 Definition of the Survival Horror Genre

In order to create a prototype to use for this study, this section will investigate the design of horror games. The horror game created for this study will be heavily inspired by the horror game Amnesia - The Dark descent since it is one of the scariest games ever made. The different elements of game design of horror games will therefore be covered and compared to how it is done in Amnesia since it successfully breaks the conventions in a number of fundamental ways. First an overall definition of the survival horror will be given.

Egenfeldt-Nielsen et. al.[5] defines the survival horror genre as games in which

"the player controls a character who has to get out of some enclosed place solving puzzles and destroying horrific monsters."

The typical emotional effect of horror is suspense which relates to worry, fear, and disgust[7]. Besides this Carroll suggests that the specific feeling of horror is invoked through disgust, particularly through monsters [2]. This fits well with the above definition where the possibility of death and the risk of not achieving the goals of the game invokes worry and fear while the monsters and setting of the game invokes disgust.
Why is horror games so effective at invoking scary emotions? Rouse argues that video games invokes fear better than non-interactive media since there is something at stake for the player[23]. In other media the fear can only be invoked in the audience through empathy with the character. In games the players project themselves into the avatar. Rouse states that this is one of the biggest advantages games have compared to other media which lead to a much more immersive experiences.

Now where the overall aspects of the genre and the feeling it invokes has been clarified the foundation for a deeper investigation of the genre can now be presented. Since video games is an emerging medium there is not much theory on how they should be designed. Film theory will therefore be used whenever can be related to video games. First the typical traits of narrative will be presented.

6.1 Narrative

Juul argues that you cannot have interactivity and narration at the same time but also admits that many games contain narrative elements[15]. One of the reasons is that actions in a story is predetermined while the essence of games is to let the player chose the action of a character. According to Kirkland this problem is often dealt with in horror games by *locating narratives in the past, by restricting interactivity, and by constructing play as a process of storytelling"*[17]. The result is that story telling in survival horror games are often more about uncovering rather than narrating or in other words playing like a detective[17]. Play therefore often involves uncovering new memories to the player as well as the protagonist[17]. The process of uncovering a story can also help the player project himself into the character since they have the same knowledge as the character they are playing.
Amnesia's narrative is a very traditional for a horror game. You play as Daniel that at the start of the game has no memory of who he is and has to recover it through diaries and letters found in the environments as depicted on figure 3.

Figure 3: A diary page the player can pick up in Amnesia

By restricting interactivity the designer can create a linear experience where the player gets guided through the game's levels and where actions has a predetermined outcome[17]. This corresponds well to the part of the definition presented earlier regarding escaping an enclosed space. This creates an experience where *"you must do X to reach Y and become a winner"*[17] which is often done by solving puzzles or killing monsters. By creating a linear game you also gain some control over space and time that you have in traditional media and you can make in-game objects and adversaries act on character movement in a scripted manner. In Amnesia this is done when the player tries to take certain routes that would lead the player away from the narrative. When the game detects this the path is blocked by red goo that Daniel cannot go through without dying as depicted on figure 4.

Figure 4: Red goo blocking the player's path in Amnesia

Finally game storytelling works best when the plot is fairly simple. A lot of nuance can then be worked into the environment and the characters the player meets. Since horror works best when as much of the narrative is left for the imagination as possible it maps well to general game storytelling[23].

To sum up this section horror games are generally linear in order to unfold a story in the right order. The stories are often simple and told by finding clues in the environment. The next section will take a deeper look at the point of view of the player along with the concept of mise-en-scene.

6.2 Mise-en-scene & Point of View

Mise-en-scene means "putting into the scene" and in films includes settings, lightning, costume, and the behaviour of the figures[1]. Geoff King and Tanya Krzywinska argue that the mise-en-scene of a game functions in much the same way as that of a film but that it also can be interpreted in more game-specific terms[16]. In many horror-based games, interiors are often splattered with gore as shown in figure 5. This does not just contribute to the background atmosphere, as would be the case in a film, but also to the players' sense of constant endangerment[16].

Figure 5: A fountain filled with blood and dead bodies in Amnesia

Horror stories are often set at locations that are recognizable to the player but which has been invaded by an evil force[23].The locations also tend to be already-familiar narrative spaces[17]. Examples are a haunted mansion, a mad scientist laboratory or as in Amnesia a castle overtaken by evil forces.

When it comes to framing of images it operates differently in films and video games due to the interactive nature of games but King and Krzywinska argues that *Any game can be analysed in terms of its point of view structure, and how this shapes the players experience.*[16].Games tends to either use a third-person perspective or a first-person perspective. When comparing the two it can be said that first-person perspective implicates the viewer in the actions and make him or her more immersed where a third-person perspective creates a distance from the actions executed by the avatar[16]. Others have also suggested that it is easier to create an immersive experience using a first-person view[36]. Amnesia also uses the first person perspective as shown on figure 6.

Figure 6: First person perspective in Amnesia

To sum up this section a first person point of view creates immersion. Mise-en-scene can be used as in film to create mood and atmosphere but can and is also used in more game related ways to influence the actions of the player. Lighting is part of mise-en-scene and since it is one of the key factors for creating a mood and invoking emotions in the player the next section will look deeper into this area.

6.3 Lighting

Lighting is considered as one of the factors along others to have an impact on the player's emotion. Niedenthal provides a set of considerations that the lighting designer must take into account[31]:

1. *Defining the basic characteristics of the environments illumination, in both time and space. Given that the player might relight the scene by adding or subtracting light (could be a flashlight) ... high-key or low-key strategies, in reference to the overall distribution of value in the scene. High-key is traditionally associated with comedy and lighter fare ... Lower-key strategies are of course appropriate for survival horror and stealth*

2. *Outlining the capabilities of the player. ... Conversely, limiting the players illumination resources, by, for example, linking flashlight usage to negative consequences ..., creates a sense of vulnerability and forces choices to be made. Once again, these decisions need to be made in relation to the game genre*

3. *Integrating illumination concerns within the whole game experience. This includes harmonizing visual capabilities of the player with other sense modalities, most importantly sound. It also includes considering the issue how to balance risk and reward, to make the choices available to the player (as actor or designer) meaningful within the game genre, according to the desired emotional complex of the game experience.*

Because fear of the dark is one of our primary frights, horror video games also render our vision partial by playing with lighting.[32]. In Amnesia this is also the case where many areas are very dark so the player is forced to light them up with tinderboxes and a lamp.

Since the emotions elicited by horror also relies a great deal on sound the next section will investigate how this is used in horror games.

6.4 Sound

As mentioned earlier King and Krzywinska also suggest that sound in games can be analysed with the use of methods developed in film studies[16]. A distinction can be made between what film studies terms diegetic and non-diegetic sound. Diegetic sound is sound that is created by the fictional world and non diegetic sound is sound that is added to the world from outside[1]. The most obvious example in games and films of non-diegetic sound is the added music that creates atmosphere[16].

Diegetic sound can both be offscreen and onscreen. Offscreen sound can create the illusion of a bigger space than what can actually be seen and provide information in a very economical manner[1]. Diegetic sound can also shape the sense of how a scene will develop which is a good tool for invoking the feeling of suspense or give the player a shock by suddenly going in another direction than what the player expects. Another function of diegetic sound is to create perspective in a scene which can be done by adjusting the level of volume and the reverberation of the sound[1].

The addition of player activity also creates new techniques for eliciting emotion beyond the audiovisual tools of films[7]. For example is goal-related evaluation of sounds an area where games differ from film. The function of game sound in goal-related evaluation is tied to the actions of the player and should provide the information necessary for decision making, thus serving a functional role for the player to reach his/hers goal. These sounds are often simple and clear in order to be helpful to the player to make it easy to distinguish between changes in game state, important events, the executing of actions etc.[7]. The ease which with a sound provides information for performing action is referred to as the functional fit[6].
In contrast to the functional fit, the narrative fit of a sound refer to how realistic the sound seems in context of the story, style and culture of the game[6]. Narrative fit is important for creating narrative comprehension[7].

Even though there is a contradiction between high narrative fit and high functional fit you can still use both. For instance a sense of uncertainty and confusion can be created by sounds that are evaluated in the most contradictory manner[6].

What is described so far are sounds which creates emotion based on the player's interest and attention but much of the emotional impact of films and games rely on the unconscious emotions[6]. It has been suggested that this is how non-diegetic music influence the perception of an event in films[26]. This concept has also been applied to sound effects in games[6].
An example of how unconscious emotions can be created is affective mimicry which is the term for the human tendency to automatically mimic emotional expressions[7]. This means that we will automatically mirror the feelings of the person we are watching or hearing without being conscious about it[7]. In Amnesia this is used when Daniel

becomses scared and starts to breathe heavily.

In this section it has become clear how diegetic and non-diegetic sound can be used to invoke emotions whether they relate to narrative comprehension, are goal related or are triggered unconsciously. It has also been clarified how offscreen diegetic sound can be used to create an illusion of a whole coherent world.

The next section will investigate the general gameplay in horror games and how game mechanics are used to invoke suspense.

6.5 Gameplay

Horror is normally used to introduce unique game mechanics based on the altered reality which is not possible in more realistic games[23]. Because of the supernatural setting the designer can throw all sorts of bizarre happenings at the player. The challenge is not to overdue this so the player feel they have lost control over the experience. So despite the supernatural setting the designer must make the game world consistent[23].

One of the main ways horror games create tension is to make the avatar vulnerable. As the name of the genre suggests the goal of the game is to survive. Vulnerability is often made by giving the evil force or monsters some advantages compared to the avatar. However the avatar's weakness should not be taken to a point where the player feels cheated by the ineptness of the character they are controlling[23]. Ways to make the character vulnerable is for instance to limit the amount of ammunition and health, make it hard to flee from the enemy, or limit the control so you for example cannot run and shoot at the same time.

Amnesia goes against the normal horror conventions in that there is no weapons or combat system in the game which only gives the player the opportunity to hide or flee. The reason for this is that it was hard for the developers to create a combat system that still makes the player vulnerable[11]. From earlier titles the developers of Amnesia learned that when players have the possibility to fight a monster they will. This can give two results. Either the player are able to kill the monster which ruins the horror experience or else the player dies and think the combat system and thereby the game is flawed. Without combat mechanics the players approached the game differently. The monsters seemed more frightening and it changed the entire experience since the combat did not steal the focus from the story and all the small things that added to the atmosphere. The developers of Amnesia also argue that you need to give the player the right tools for the task in order to inform the players which role they are playing[11]. For instance in Amnesia the player has to uncover a mystery so the player is given an oil lamp rather than a gun.

Another popular way to create tension is to give the player some information and leave some out. This can for instance be done by noticing the player about an enemy but what kind of enemy and exactly where it is, is unknown which creates a tense experience. This can be done by diegetic sounds coming from dark or foggy areas[23].
The whole concept about limiting the players view with fog, darkness or fixed camera angels is commonly used in horror games. The reason for this is that if the vision is partial it creates a feeling that the enemy could be everywhere which again emphasises the vulnerability of the player[32]. Like the case with point of view explained in section 6.2 this restriction of view can be used to invoke different emotions.

A common way to give the monsters an advantage and argue for the avatar's weakness is to use a panic/insanity meter which is used to limit the control, decrease the vision and confuse the player[32]. This is also an area where Amnesia does something different. It contains none of what the developers call "competitive mechanics" which means that the game adjusts to the player rather than than forcing the player to practice to beat it[11]. This was done by eliminating the penalty of death and use enemies and sanity as atmosphere rather than competitive mechanics. In other games it is also fatal if the player's sanity is reduced to zero. The developers of Amnesia discovered that it was hard to balance the sanity mechanic and found it much more effective to just keep it as an effect to add atmosphere to the game. The AI of the monsters in Amnesia is also adjusted to increase the atmosphere of the game. An example of this can be seen on figure 7 where the player is hiding from a monster. If the player is hiding, the monster moves close to the player's position and begins sniffing and looking for the player before it goes away. This builds tension and really adds to the atmosphere.

Figure 7: Screenshot from Amnesia where the player is hiding from a monster

This section has clarified the gameplay of horror games and how it is used to invoke emotions in the player. Game mechanics and elements that are common in horror games has also been introduced that could be used for this project.

7 The Prototype

7.1 The Game

In this section we will be describe how and why the game prototype is designed and implemented as it is. Maya, Blender and Unity was used to implement the game, as they seemed to fit the scope of this project.

The story of the game is typical for horror games. The player takes the role of a person who wakes up in a strange location without any memories of why he is here, the player then has to escape this location. As stated in section 6.1, the locations should be familiar, based on this a hospital setting was chosen since it is a familiar place for many people. More specifically, the location is an underground hospital inspired environment. The underground setting was chosen so that there would not be any windows. This gives the player a feeling of enclosure and isolation which is an important part of the survival horror genre as stated in section 6. This setting also have the added advantage of being easy to model and texturize.

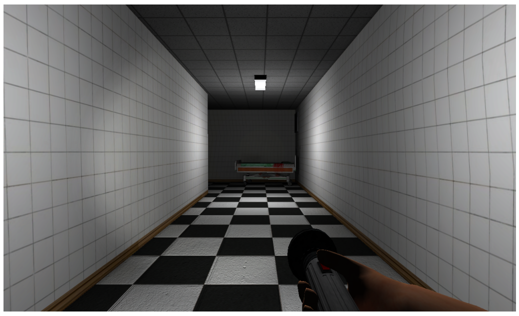


Figure 8: In game screen shot of the underground hospital inspired environment

Based on the criteria above, it was decided to have the following storyline: The player is being transported to an operation room. He is disorientated and confused. During the operation something supernatural attacks the hospital leaving it deserted and destroyed. When the player wakes up he can't remember where he is or why he is there. The objective of the game is then to escape the hospital.

7.2 Gameplay

This section describes the choices that went into the design of the gameplay. The first choice was to not allow the player to be able to fight the monsters. This is based on the design principles from section 6.5. The only chance the player has to survive is to flee from the monster and hide. This should make the player more aware of changes in the surroundings and therefore enhance the feeling of tension.

Figure 9: Different camera types. Left: Silent Hill, third person. Right: Amnesia, first person.

Another choice that had to be made was if the game should be third person view or first person view. A common camera angle in the survival horror genre is third person. This is seen in the Silent Hill games and allows for the camera to be placed more like the camera would be in a horror movie. On the other hand choosing a first person point of view creates more immersion, and at the same time restrain the players options view wise, as stated in section 6.2. It will also avoid the possibility of creating a distance to the protagonist that would occur if it was third person. Since creating emotions are one of the key points when designing a horror game it is chosen to have a first person view for the game because of the increased degree of immersion it gives. Another key point, from section 6.2, is that limiting the player's view makes them feel more vulnerable.

We also had to consider the capabilities of the player respectively to the genre. We decided that the player should be equipped with a flash light. The flash light serves the obvious role of lighting up the environment in front of the player so he or she can navigate the level. To make the flash light a bigger part of the gameplay, it was made so that when it is used, it would run out of battery over time. This forces the player to think about when to use or not use the flash light, creating a feeling of vulnerability, because one does not want to run out of power. Another role that the flash light also plays is that the player has to use it to avoid becoming insane. The insanity system was implemented in the same way as it is in Amnesia where the player slowly becomes insane when standing in darkness. The insanity system has no particularly gameplay role at the moment, other than to stress the player.

The considerations suggested in section 6.3 were in mind when the game was lit. First consideration was to determine the overall lighting of the environment in time and space. Also keeping in mind whether or not the player has influence of the overall light, and lastly which strategy is used. The first part of the game where lit with a medium amount of light because the player will not be equipped with the flash light from the beginning. In level two this changes when the player gets the flashlight. The lighting is therefore changed to lower-key lighting right before achieving the flashlight so the player appreciates the acquisition of it. After the player has received control of a light source, lower-key lighting is used in all the levels afterwards, as it is the most common strategy used in the horror-genre. Though the last section of level six is lit up so they can feel somehow relaxed.

Figure 10: The last section of level six

The game is designed so there is approximately 10-15 minutes of gameplay since this is a common interval for game studies[25].

7.3 The Levels

Because the goal is to measure the physiological response from the player when experiencing scary events in a horror setting we also need to have some neutral events to compare the data against. It is chosen to make a total of six levels where we have 3 calm levels, which is levels where no events occur, and 3 levels with scary events. Level one, four and six are the calm levels. Two, three and five are levels with scary events. It is done this way so there is no obvious pattern in the sequence of scary levels.

Level one and two work as tutorial levels. In level three the player will encounter a monster. Level four is a yet again a calm level, though because of the lower-key lighting it still is a little unnerving, but they do not encounter any scary events. In level five there is one puzzle and one scary event, and level six, which is the last level, is also the last calm level.

Level one is a tutorial level. Here the player will be introduced to the basic controls. It is a calm level since the player has not yet learned of to navigate in the game. **Level two** is also part of the tutorial. Here the player is introduced to the flashlight and the insanity system. They are introduced to the insanity system by walking down a hallway when the lights suddenly turns of, this makes level two one of the scary levels. **Level three** is also one of the scary levels. Here the player is introduced to a monster. As he is walking down a hallway the double doors at the end of the hallway is smashed open and the monster appears. The player then has to flee and hide from the monster.

Figure 11: The encounter with the enemy in level03

Level four is then a calm level. The player has to navigate through a series of hallways and rooms like a small labyrinth. **Level five** is the last scary level. Here the player sees a moving wheel chair at the beginning of the level, when the player goes after the wheelchair he will end up in a hallway that keeps looping around in an infinite circle. Only when the player changes direction and starts to walk back will he get out of the loop. Here he will face a long dark hallway and when he moves forward the wheelchair from before will be lunched out from the dark hallway towards the player. **Level six** is the last level, and is also a calm level. In level six the player finds an elevator but because it's broken he will need to find an alternative way up. Level six ends when the player finds some fire stairs in the back that leads to ground level so he can escape the hospital.

7.4 The Six Events

For the test there will be a total of 6 events throughout the game that the test participants should remember when they have played the game (See section 8 for more information how this self-reporting is designed). The 6 events will be spread out over the 6 levels so there will be one in each. The following is a description of the events:

- Event 1: At one point during the tutorial the player is asked to jump over a bed.
- Event 2: Here the player is moving down a hallway when the lights begin to go out.
- Event 3: The player will encounter the monster. It will break through a double door in front of him.
- Event 4: The player will walk by a battery that he can pick up.
- Event 5: The wheelchair will come towards the player at high speed.
- Event 6: Right before the player finish the game, he will walk down a long hallway that are lit by windows.

Figure 12: Event 5, the wheelchair will come towards the player at high speed

Each event only happens once during the game so that the player will not confuse them with other events.

7.5 Sound Design

The sound design in the prototype is based on the theory from section 6.4. The sound is primary found on the internet and recorded with a

ZOOM H4 recorder and edited with the Digital Audio Workstation Reason by Propellerhead. Some of the non-diegetic sound effects for creating atmosphere, low level emotional reaction and building up tension were programmed using the built in digital synthesizers in Reason. The balance between narrative fit and functional fit for the goal related sound in the game was made so they did not seem too distinctive from the rest of the game world but so they still were functional. The reason for this design choice was to avoid loss of immersion so the players more easily could project themselves into the game world as it would seem coherent. To increase the narrative fit of the diegetic sound, Unity's build in reverb filter was used to create reverberations that fit the different locations. Unity's 3D audio engine was also used on the diegetic sound so the player would be able to locate sound in the game. Both those coming from actually game objects and those from imaginary offscreen events. As mentioned before the prototype consist of two types of levels, some with calm event and some with fear invoking events. The levels with calm events don't have other sounds than some offscreen background noise like vents to make the game world seem alive and the goal related sounds triggered by the player's interaction. The levels with the fear invoking events use both diegetic and non-diegetic sound to invoke low level emotions and build up tension and player expectation.

7.6 Limitations

As stated in section 2.3, the IOM-device is used to measure the the physiological data from the player. It has been noted that using the device together with a mouse can be difficult because you need to wear one of the sensors on your index finger, the same finger that is used when clicking the mouse as shown on figure 13. As mentioned, Biofeed the Zombies used this setup and concluded that it might have influenced their test results.

Figure 13: How to use the device with a mouse

A solution to this problem would be to interact with the game through a controller. We chose a xbox 360 controller, with this the player could wear the sensors as shown in figure 14. This setup allows the player to wear the heart rate sensor on one hand and the galvanic skin response sensors on the other hand, leaving the thumps and index fingers free to interact with the controller as normal.

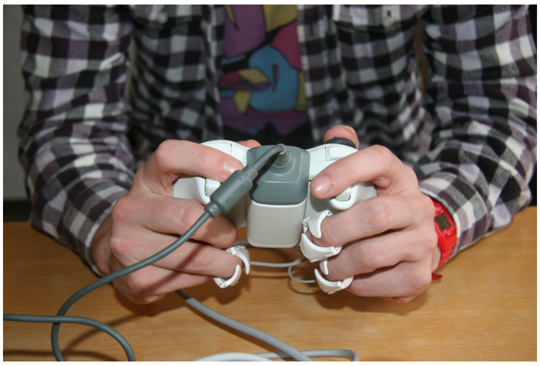


Figure 14: How to use the device with a controller

We tried to map the controls so that they would be as intuitive as possible, this is done to resemble other Xbox 360 games' controls. Face-button A is used for jumping, face-button B is used for turning on/off the flashlight, and face-button X is used for putting in new batteries in the flashlight. The last face-button Y is used to check how many batteries the player has in his inventory. Right trigger is used to open/close doors. Left stick is used to move and the right stick is used to look around.

8 Test

This section will excplain the design of the final test.

8.1 Test Method

The test was designed as a quasi experiment. A quasi experiment is categorized by being similar to normal experiments but without random assignment of conditions [14]. This presents a risk of low internal validity, meaning how valid it is that two variables are causal. It can also lead to confounding variables - extraneous variables that correlate both with the independent variable and the dependent variables. However since the content of the test is to play a game prototype in a controlled lab condition it is unlikely that there will be confounding variables.

Although the order of levels in the game prototype could have been randomized it would have been harder to plan for calm and scary moments this way. Since there was a low risk for the non-randomization of events to produce a biased result, it was chosen to do the test as a quasi experiment.

8.2 Test Procedure

The procedure of the test is as follows. First the participant was asked to read and sign a consent form that also explained that the experiment requires the recording of their physiological signals. Afterwards a baseline reading of the participants physiological signals are recorded. In order to reduce noise in the physiological signals the participants were asked to rest their hands on a table and keep them steady during the experiment. During a read through of related literature it was also not possible to find out for how long time a baseline reading should be taken. We settled on 2 minutes since a pretest showed that a longer period frustrated the participants and could create a bias in the baseline reading. When the baseline had been recorded the participant played through the game.

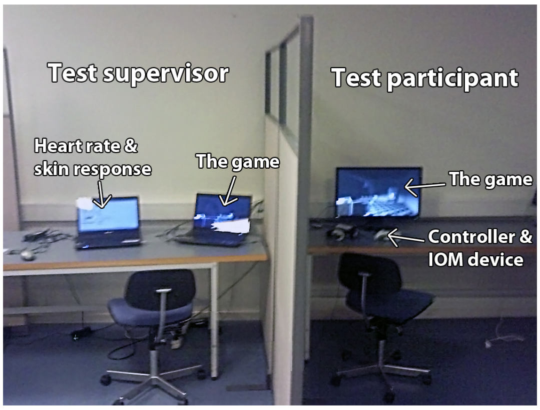


Figure 15: Test set up

During the play through the test supervisor sat and monitored the data received from the IOM device and follow test participant's progress in the game.

Afterwards the participants filled out a questionnaire regarding their emotional state during the events described in the section on the game prototype as well as demographic questions to record other extraneous variables. This was done at the end since other studies have determined that interaction with researchers and the act of filling out questionnaires affect the physiological state of the test participants[25, 24]. The following section describes the design of this questionnaire.

8.3 Design of Questionnaire

As explained in the section regarding emotions it is chosen to use the PANAS-X scale albeit in a modified version.

The modified version contained only the negative affect scale and the fear scale. Only these were used since these were the only dimensions that were needed to determine the amount of fear to each event. This was done to lower the fatigue of filling out the questionnaire. To further prevent fatigue the test participants was told they could choose not to rate a word and it would be interpreted as if they had rated that word with a 1.

In order to help the test participant remember each event in the game a screenshot and a description of the event were provided. The participants also had the option of providing additional words to describe their affective state.

To take extraneous variables into account the questionnaire also asked the participants about their gender, age, how much they like to play horror games, their overall experience with playing video games and how experienced they were with playing games with a game pad.

9 Results

22 participants participated in the test. Participant one and 16 is not included in the results since they did not finish the game. This leaves a sample size of 20 test subjects. The results presented in this section has been chosen because it is the ones with the biggest contrast between calm and scary events. The rest of the results can be seen in appendix D and the raw data can be found on the CD. Firstly the results of what the participants self-reported are shown. This is followed by a description of the data collected by measuring the participants galvanic skin response and this is again followed by a section describing the heart rate results. Lastly, the results of comparing the self-reported answers and GSR are showed.

9.1 Self-reported results

Figure 16 and 17 describes two of the events. Event 5 tries to cause the participants to feel fear, whereas the other is trying to calm them down.

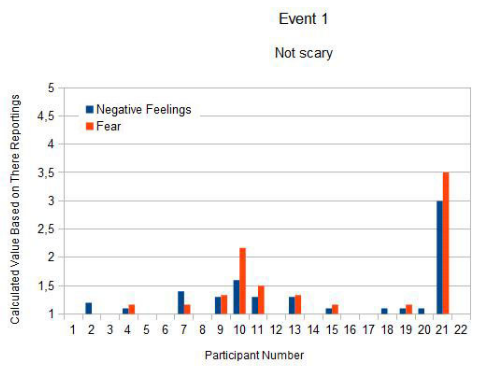


Figure 16: Emotional responds for event 1

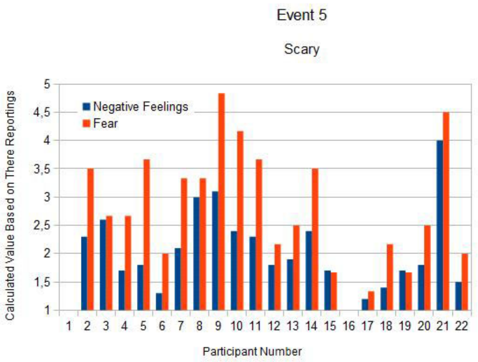


Figure 17: Emotional responds for event 5

The x-axis shows the participant number and the y-axis the feeling using the PANAS scale. For each event the emotion is calculated as the average of the reported values for the adjectives related to that emotion. Fear is calculated based on six adjectives and is shown in red. Negative feeling consists of ten adjectives and shown in blue. When comparing the diagrams for the two events, one can immediately see that the negative feeling and fear reported in event 5 are higher than the reported values for event 1. There are though deviations from this pattern. Participants 21 in event 1, felt fear in both events and for participant 17 the responds is miniscule.

9.2 GSR Data Results

The galvanic skin response was measured for each participant through the whole game. Additionally each participant got a base line measured for two minutes before playing the game. Videos were here after rendered with game screen capture of the participant playing and GSR side by side. The videos were studied to find correlations between events in the game and the participant's level of GSR. 18 describes the average skin responds of all participants for two events.

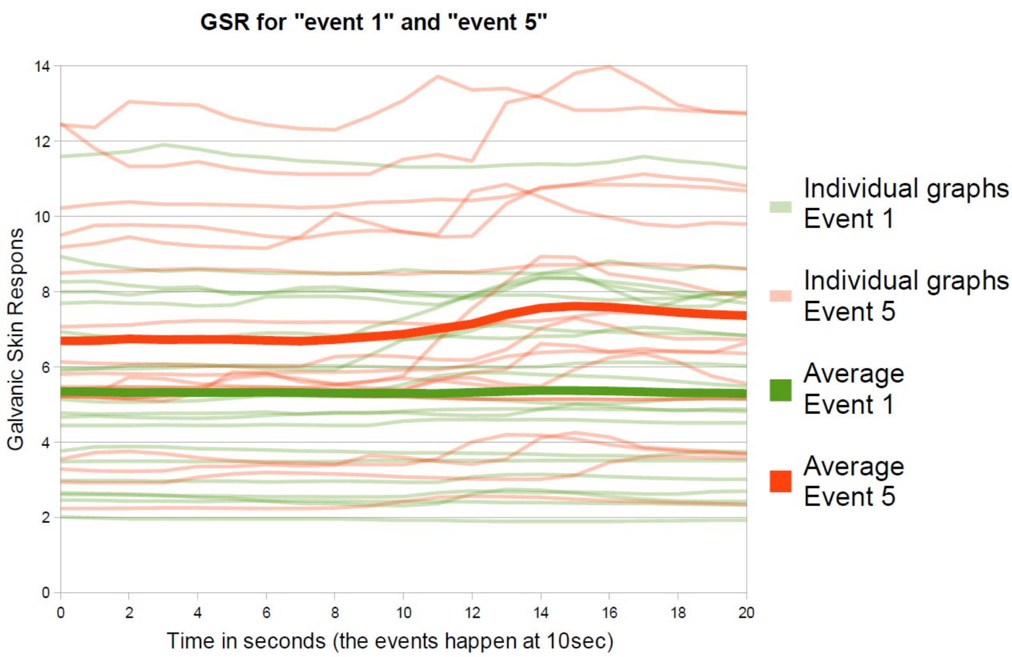


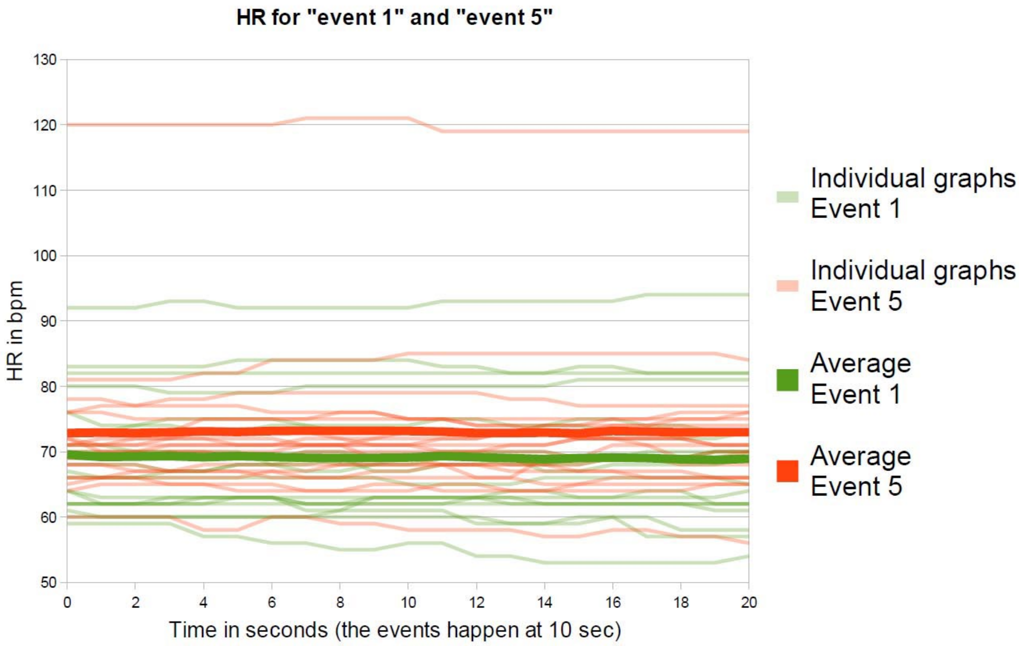
Figure 18: Average GSR curve

The x-axis denotes time in seconds. The specific events happens at x = 10 seconds. The y-axis shows the GSR data received by the biofeedback device. Event 1 is supposed to be a not scary event and is denoted by the green graph and event 5 is described by the red graph. The opaque graphs are the individual skin responds of each participant.

As can be seen in the figure, the average GSR is increasing in the scary event whereas remaining steady in the non-scary event. Notice the opaque graphs, showing how much physiological responds vary from person to person. Also notice the difference in the participant's base line. There seems though to be areas of concentration. The participants respond to the non-scary event (the green opaque lines) seems to be concentrated between 2 and 5, whereas their respond to the scary event seems to be concentrated between 5 and 8.

9.3 Heart Rate Results

Figure 19 describes the average heart rate in BPM for the same events as in Figure 18. Despite that in individual cases there can seem to be a responds in the heart, this could just as well be random occurrences. The overall picture is therefore that there are no informational fluctuations in the measuring heart rate.



2

Figure 19: Average BPM curve

9.4 Relation Between Fear and GSR

In figure 20 below the overall relationship between the reported fear and the GSR data is visualized. The blue squares denote one event for one participant. The x-axis describes the normalized GSR. GSR is normalized using the lowest change in GSR as 0 and the highest change as 1. The y-axis describes the normalized reported fear. This is done by normalizing the calculated fear value.

Notice that for this method of relating GSR to fear had to show a relation the data points must be concentrated around the diagonal going from origin. In this result the data point's deviates for much from the diagonal and does therefore not show any consistent relationship between the magnitude of reported fear and the physiological responds in GSR.

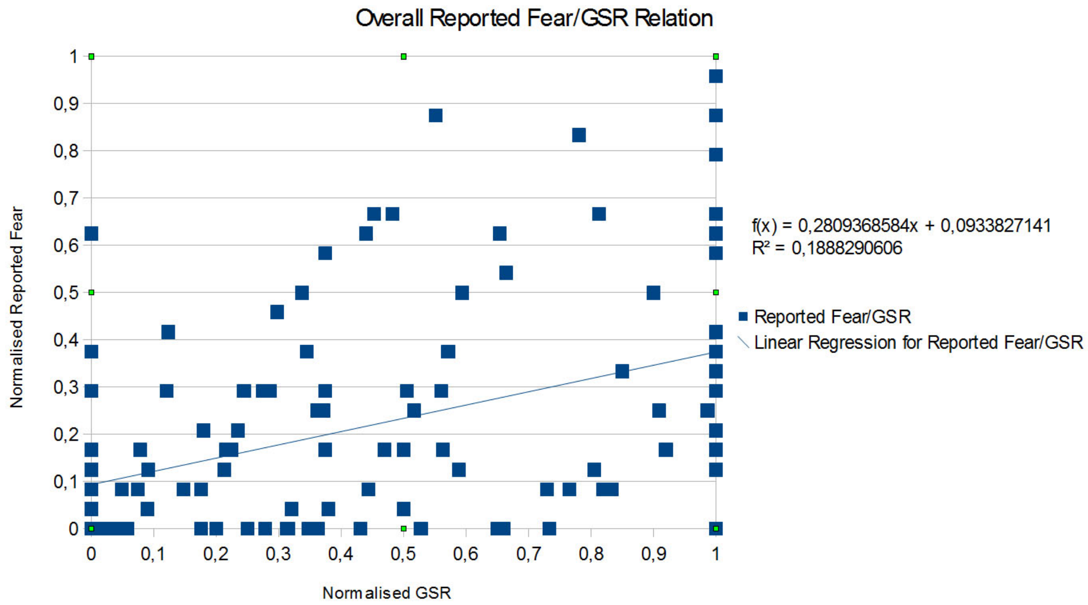


Figure 20: Overall relationship between GSR and the reported fear

9.5 Comments Reported

Some of the participants reported that they where annoyed at some points in the game mainly due to technical issues. For instance a participant reported the following comment at the first event *"The steering annoyed me a bit, it took a while to get used to"*. A couple of participant were also expecting things to happen in the calm places. For example participant 22 makes the following comment to event 4 *"Mostly because I, in general this far into the game was expecting something nasty to happen"*. Missing adjectives on the questionnaire was also a common comment participant 12 writes to event 5 *"Shocked = 5 startled = 5"*. There were also comments on the overall experience, some with feedback for what could have made the game better and other just complimenting the game. Another thing that should be noted is that it seemed like there were a bit of confusion about the meaning of the adjectives on the questionnaire since some participants asked while they were filling out the questionnaire.

The next section will be a discussion and analysis of the result presented.

10 Discussion

Overall emotions are very complex and have an extremely subjective nature. There does not exist an approved errorless way of transferring them into data. The validity of the results will therefore always suffer from this regardless of the choice of measuring framework. This must be taken into account when reviewing and analysing the data.

It was impossible to find a general correlation between GSR and self-reported fear. A linear regression line fitted the data the best but with an R-value of 0.19 the result is far from valid. The subjectiveness of the participants responses in both GSR, HR and self-reported fear is also immediately seen when reviewing the results. This makes it hard to process the data since it is necessary to adjust the processing method to each participant. This indicates a need for an intelligent system which can be calibrated so it suites the physiological response of each person. This individuality in the physiological response influences all of the results.

Each time the level of GSR is increased it does not return to the same level as before the increase. This means that GSR of each participant seems to be increasing continuously throughout the game. How much it increases is also highly individual. It seems to be the game influencing the increase, since no increase happened while not playing. Several participants reported the feeling of fear in the situations meant to be calm. This could be the reason for the steady rising GSR of the participants. If this is not the case the signal must be filtered accordingly if one is to use GSR for affective horror games.

Another important point is that the GSR did not increase solely due to the scary events in the game. A lot of not scripted events caused fluctuations in the GSR signal. This allude to the fact that other emotions are influencing the GSR of the participants. Despite that the biggest fluctuations is caused by the scary events, these other emotions could cause interference. Measuring across an additional modality in order to filter out these emotions, could become inevitable in real affective version of the prototype.

The BPM of the participants did not really correspond to any of the events in the game. This contradicts what Segal and Dietz [35] concluded in their research, where they tested participants playing Ms. Pacman. Despite this, there seems to be to many issues in regards to use BPM data of participants for affective gaming.

In regards to the self-reported results, there seems to exist consistency between the reported level of fear and the events happening in the game. Despite, it must be taken into account that people are not good at self-reporting their emotions [25]. This is well-known and PANAS tries to take this into account by providing multiple words to describe the same emotion. Despite the fact that the participants do not report their emotions real time influences the results. Firstly the participants can have a hard time remembering what they felt. If this is the case they are either influenced by the expression of the pictures shown, guessing or just reporting that they felt nothing. Both of these problems have properly contributed to the randomness of our results.
Secondly, at the time of reporting the participant have played through the whole game. This means that their reported emotion is influences by the ratio between the other events in the game, instead of the actual emotion they felt at the specific event. This causes an event, a participant experienced as really scary, to be reported as less scary, because the participant compares it to scarier events.

Another problem became apparent when one test participants, reported that he was having problems with understanding some of the English adjectives describing emotions in the questionnaire. This could unfortunately indicate that others might have experienced the same problem, but didn't express it. This means that the relationship between how much fear and negative feelings the participant felt might be wrong.

In regards to the relationship between GSR and fear, the current way of calculating it is far from flawless. One problem relates to the way the data is normalized. The normalization is currently based on the biggest change in a participants GSR data. This means the data of a person with only small changes is GSR can end up identical to a person with large changes. Since physiological responses are very depend on the person, it is necessary to normalize the data individually. Even so it leaves the possibility that a person, not responding physiologically, will end up influencing the final results as if he had high response.

Another problem about the comparison of physiological data to the self-reporting data is their different sample rate. The GSR data is one sample per seconds, where the self-reported fear has a sample rate of 6 samples per game (approx. 12 min). This could indicate that other means of self-reporting is needed if one has to compare these two kinds of data.

Additionally, despite the contrast between calm and scary moments being a tool for investigating the relationship between fear and arousal, the questionnaire does not directly try to measure how calm the participants felt. It was therefore not possible to pin point the difference in emotional states of the participant between calm and scary events. This is instead derived as the difference between the person feeling fear or negative feelings and not feeling any of these. This could possible cause the results to be less accurate.

11 Conclusion

Through this study it was discovered that the precise correlation between physiological arousal and magnitude of self perceived fear had not been researched before. It was therefore decided to conduct a test to determine this correlation. It was impossible to see any clear relation between GSR and self-reported fear when looking at the processed data from each measurement. A linear regression fitted the overall data the best but the validity of the result is too low to know if it is the true tendency. This could indicate that the processing of the raw data is not made properly. The chosen way of self-reporting emotions was also found to be an inaccurate measurement, since some participants could not understand the adjectives in the PANAS-X scale used for the test. This study shows that another scale than the Affect Grid or PANAS-X is needed to do quick measures of specific emotions such as fear and frustration.

Since an increase in GSR and self-reported fear during the scary events was observed there seems to be a relation between arousal and fear. However due the difference in the way GSR increases between the test participants it was not possible to conclude on any universal relation between GSR and fear. One way to solve this could be to have a profile for each player that describes how the player's physiological data changes in different situations.

Our test did not show any relation between heart rate and the self reported fear of the player. As mentioned in the discussion this is most likely because of the chosen sensor and the filtering used. Furthermore it seemed that trying to calculate BPM in real time using the last five beats was inaccurate and that using BPM calculated over a full minute is more precise. It might also give a better way to measure arousal over a longer period of time then compared to GSR that seems to be more instant.
Lastly, both BPM and GSR seemed to change randomly during game play at places where it should not change. So it is not possible to use that alone to estimate how afraid the player is. It is therefore necessary to implement some kind of artificial intelligence that can assess if the change is caused by something scary happening in the game or not.

12 Future Work

12.1 Test method & Game Prototype

In order to obtain better results from the test of the prototype, several areas can be improved. The biggest area to improve is how to assess emotions of the participant. As mentioned in the conclusion the PANAS scale was found to take to long time to fill out. One solution could be to accept that there will be a bias from frustration in the arousal levels and omit the negative affect dimension from the questionnaire. Another possibility which likely will lead to better results, but be more time consuming would be to create an entirely new affect scale more suited to measure categories of emotions. Such a scale would also be beneficial for the general affect research community in cases where the Affect Grid is not sufficient.

Another approach is to measure more modalities. The most obvious is facial expressions. However as reported earlier video recordings of the participants is likely to yield inconclusive results. A better approach would be to use electromyography to measure directly on the participants facial muscles. Research has shown that facial muscles are activated when emotions are felt even though they are not accompanied by a facial expression that can be seen. It could also be beneficial to look at other available data from the IOM device such as heart rate variance and skin conductance response rate that have been shown to have correlation to fear [20].
The quality of the game prototype could also be improved so there is a stronger difference between calm and scary events. Futhermore the interaction issues discovered during testing should also be fixed.
The IOM device should also be replaced with a device that is less intrusive. This would also mean the game could use a traditional mouse and keyboard control scheme which the first person perspective is better suited for.

12.2 Reflections on Affective Gaming & emotions

From the experiences gained through conducting the research for this work, implementing the physiological measurement device and analysing the data, we have come to the conclusion that there have to be made large strides in the areas of obtaining emotions from physiological data and self reporting in affective research. The lack of another tool than the Affect Grid means that it is hard to conduct research such as the one described in this journal.
The approach to derive the player's emotions from physiological data also needs to be very accurate. Otherwise the player will become frustrated. On top of that, even if the approach can identify the emotion the player is having with 100% accuracy, it could still be a frustrating experience since the player might not know what his or her own feeling is at the moment.

Emotions are furthermore so complex that it is still uncertain if it is possible to describe their nature. Even more uncertain is whether they can be described in a way computers can understand. Only further research into affective computing can give a definite answer for this.

12.3 Proposal for an affective horror game

As can be seen from this research the easiest affective parameter to measure is arousal since it can be obtained directly through galvanic skin response. The tricky part is to asses the affective state of the

player. Instead of using time and money on implementing additional sensors in order to measure more modalities, it might be a better approach to measure arousal and validate this by analysing the situation in the game. For example if the player sees a monster and arousal goes up it is likely to be a result of fear. On the other hand if the player keeps dying at the same place in the game many times in a row and the amount of arousal after each death increases for each time, it is likely to be connected to frustration caused by dying. This would require extensive play testing in order to find out which emotions the events in the game triggers.
As found out in section 2.2 the adjustments to the game should be made to background variables. In a horror game this refers to whether or not a scary event should be triggered or how many monsters there should be created when the player enters a new location.

References

[1] D. Bordwell and K. Thompson, *Film art: an introduction*, McGraw-Hill, 2010.

[2] N. Carroll, *The nature of horror*, (1987).

[3] R. J. H. Christopher P. Barlett and C. Bruey, *The effect of the amount of blood in a violent video game on aggression, hostility, and arousal*, (2008).

[4] A. Dekker and E. Champion, *Please biofeed the zombies: Enhancing the gameplay and display of a horror game using biofeedback*, The University of Tokyo, 2007.

[5] S. Egenfeldt-Nielsen, J. H. Smith, and S. P. Tosca, *Understanding video games: the essential introduction*, Taylor & Francis, 2008.

[6] I. Ekman, *Psychologically motivated techniques for emotional sound in computer games*, (2008).

[7] I. Ekman and P. Lankoski, *Hair-raising entertainment: Emotions, sound, and structure in silent hill 2 and fatal frame*, McFarland & Company, Inc., 2009.

[8] P. Ekman, *Basic emotions*, John Wiley & Sons, 1997.

[9] K. M. Gilleade, A. Dix, and J. Allanson, *Affective videogames and modes of affective gaming: Assist me, challenge me, emote me*, University of Vancouver, 2005.

[10] M. D. GRIFFITHS and I. DANCASTER, *The effect of type a personality on physiological arousal while playing computer games*, (1995).

[11] T. Grip, *Evoking emotions and achieving success by breaking all the rules*, August 2011.

[12] R. L. Hazlett and J. Benedek, *Measuring emotional valence to understand the user's experience of software*, (2007).

[13] E. Hudlicka, *Affective computing for game design*, McGill University, Montreal, Canada, 2008.

[14] S. Jackson, *Research methods and statistics : a critical thinking approach / Sherri L. Jackson*, Wadsworth/Thomson Learning, Belmont, Calif. :, 2003.

[15] J. Juul, *Games telling stories?- a brief note on games and narravtives*, (2001).

[16] G. King and T. Krzywinska, *Film studies and digital games*, SAGE Publications, 2009.

[17] E. Kirkland, *Storytelling in survival horror video games*, McFarland & Company, Inc., 2009.

[18] M. Kivikangas, I. Ekman, G. Chanel, S. Jarvela, B. Cowley, M. Salminen, P. Henttonen, and N. Ravaja, *Review on psychophysiological methods in game research*, University of Stockholm, 2010.

[19] I. Knez and S. Hygge, *The circumplex structure of affect: a swedish version.*, (2001).

[20] S. D. Kreibig, F. H. Wilhelm, W. T. Roth, and J. J. Gross, *Cardiovascular, electrodermal, and respiratory response patterns to fear and sadness-inducing films*, (2007).

[21] K. Kuikkaniemi, T. Laitinen, M. Turpeinen, T. Saari, I. Kosunen, and N. Ravaja, *The influence of implicit and explicit biofeedback in first-person shooter games*, ACM, 2010.

[22] P. Lang, *The emotion probe: Studies of motivation and attention*, (1995).

[23] R. R. lll, *Match made in hell: The inevitable success of the horror genre in video games*, McFarland & Company, Inc., 2009.

[24] R. Mandryk, K. Inkpen, and T. Calvert, *Using psychophysiological techniques to measure user experience with entertainment technologies*, (2006).

[25] R. L. Mandryk and M. S. Atkins, *A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies*, (2007).

[26] S. K. Marshall and A. J. Cohen, *Effects of musical soundtracks on attitudes toward animated geometric figures*, (1988).

[27] Metacritic.com, December 2011.

[28] C. H. Mohammad Obaid and M. Billinghurst, *"feed the fish": An affect-aware game*, ACM, 2008.

[29] L. Nacke, M. Kalyn, C. Lough, and R. Mandryk, *Biofeedback game design: Using direct and indirect physiological control to enhance game interaction*, 2011.

[30] L. Nacke and C. A. Lindley, *Flow and immersion in first-person shooters: measuring the player's gameplay experience*, New York, NY, USA, 2008, ACM.

[31] S. Niedenthal, *Complicated shadows: The aesthetic significance of simulated illumination in digital games*, (2008).

[32] B. Perron, *Coming to play at frightening yourself :welcome to the world of horror video games*, (2011).

[33] J. A. Russell, *A circumplex model of affect*, (1980).

[34] J. A. Russell, A. Weiss, and G. A. Mendelsohn, *Affect grid: A single-item scale of pleasure and arousal*, (1989).

[35] K. R. Segal and W. H. Dietz, *Physiologic responses to playing a video game*, (1991).

[36] B. Strauss, *Be yourself (or not): Bioshock*, February 2011.

[37] D. B. Tim J. W. Tijs and W. A. IJsselsteijn, *Dynamic game balancing by recognizing affect*, Springer-Verlag Berlin Heidelberg, 2008.

[38] D. Watson and L. A. Clark, *The panas-x: Manual for the positive and negative affect schedule - expanded form*, (1994).

[39] D. Watson, L. A. Clark, and A. Tellegen, *Development and validation of brief measures of positive and negative affect: the panas scales.*, (1988).

[40] F. Yasuma and J.-i. Hayano, *Respiratory Sinus Arrhythmia : Why does the Heartbeat Synchronize With Respiratory Rythm?*, American College of Chest Physicians, 2004.

Appendices

A Ekman's basic emotions

The strongest proponent of a categorical definition of emotions are psychologist Paul Ekman. Ekman's framework is based on Charles Darwin and Tomkins[8]. He views emotions as having evolved as something that gave the organism a better chance for survaival. Based on studies on facial expressions he also concluded that emotions are to complex to be represented by arousal and valence, the two most often used dimensions in affect research. Instead Ekman has made a definition for what a basic emotion is. With basic emotion is meant an affective response to a current fundamental task that helps the individual adapt to it. The most relevant parts of this definition in relation to affective gaming is:

universal distinctive signals and prior events
: Emotions act as a way to signal to other individuals what the individual's emotional state is. This is often through multiple modalities such as context of speech, facial expressions, and body language. Certain emotions are also linked to certain type of prior events. For example is fear related to events where the individual is at risk of suffering physical or psychological harm.

Emotion specific physiology
: emotions prepares the body to the situation that created the feeling. For example is fear connected to the flight-fight mechanism.

Quick onset & brief duration
: Emotions are quick to emerge and has a brief duration, contrary to what Ekman defines as moods that are longer in duration.

automatic appraisal
: The individual experiencing an emotion often does so automatically, it does not require a concious effort.

Unbidden occurrence (involuntary)
: What emotions we feel is not something that can be decided upon consciously.

Subjective associations & experiences
: For each individual an emotion is linked to distinctive thoughts, memories, and images experienced by that person. This also means that each basic emotion is experienced differently across individuals.

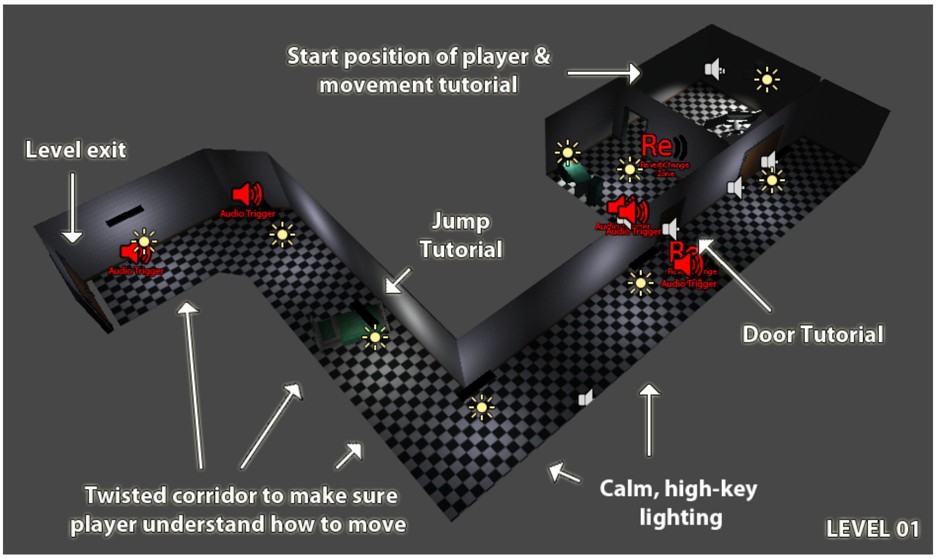
The last part of the definition that is not so relevant for this work is:

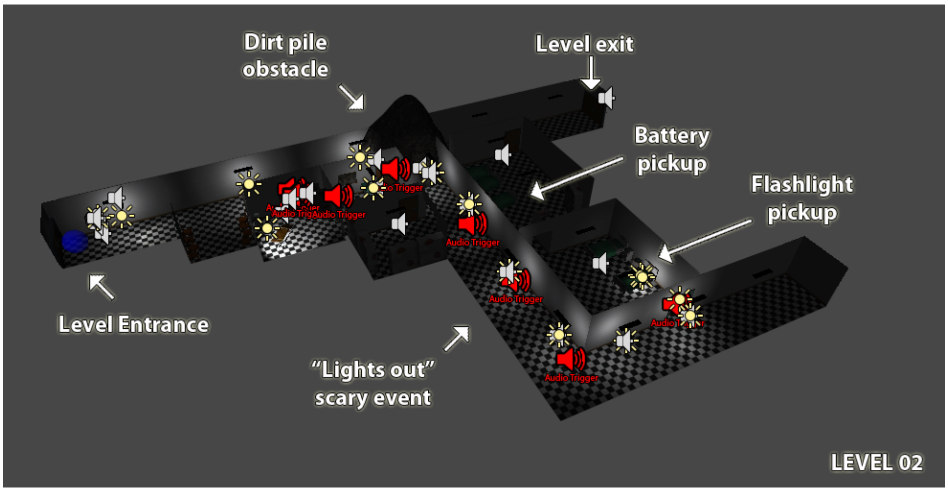
distinctive appearance developmentally
: The age of the individual seems to dictate what types of emotion the individual can express.

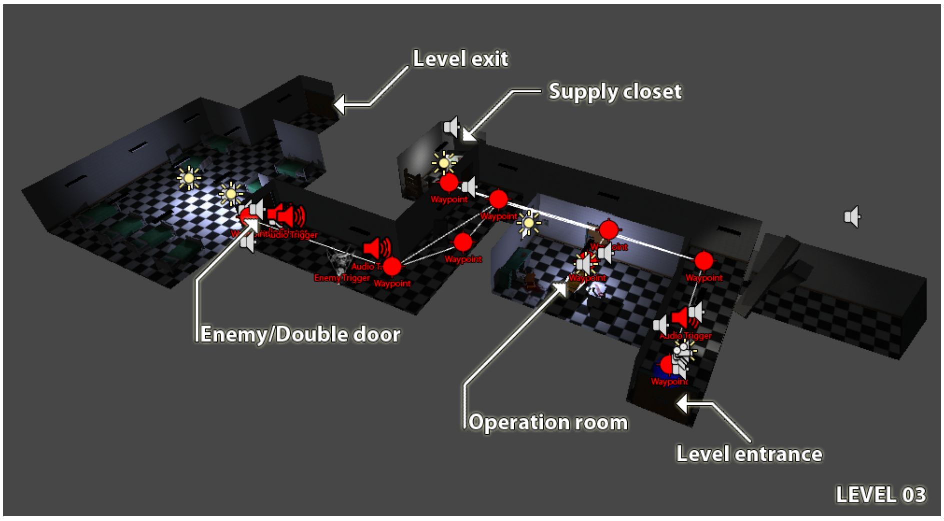
presence in other primates
: The emotion needs to be present in other primates in order to be basic since basic emotions are are a survival tool created by evolution.

Based on this definition Ekman created his own categorization of emotions denoted emotion families[8]. A family of emotion consists of emotions that share what Ekman calls a common theme, a set of characteristics similar to all the emotions in the family. It is interesting to note that fear denotes its own family while frustration does not have its own family but could belong to both the families of disgust, contempt, and distress. This could indicate that frustration is a more complex emotion than fear. It is chosen to use Ekman's definition of emotion since it lends itself very well to research into affective gaming since it takes into account the physiological aspect of emotions and gives description of the duration and onset of emotions.

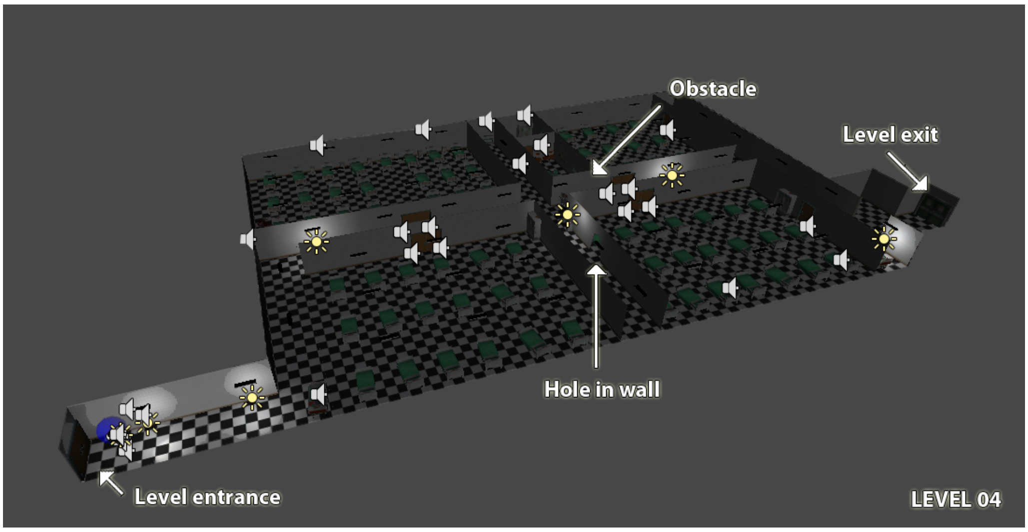
B Level 1-3 overviews

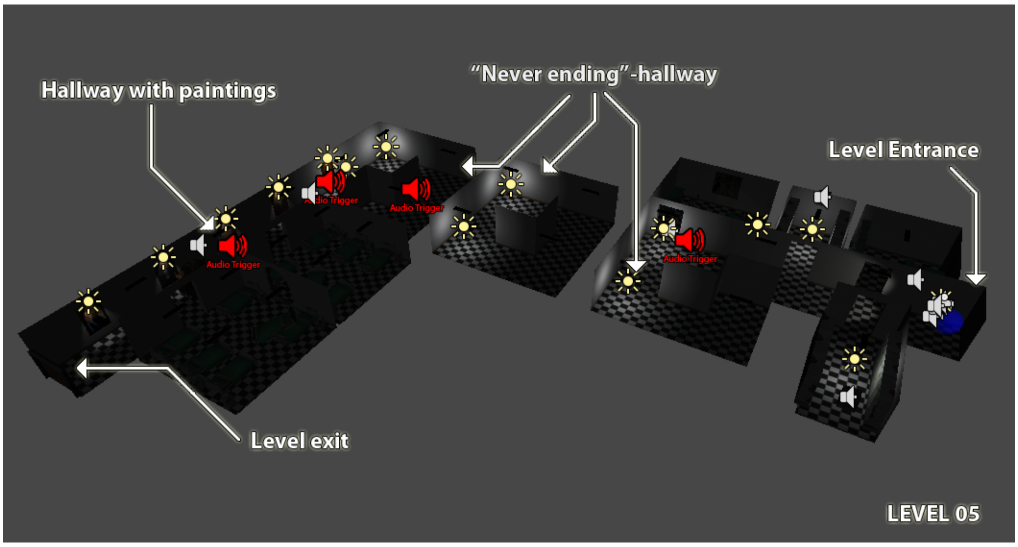


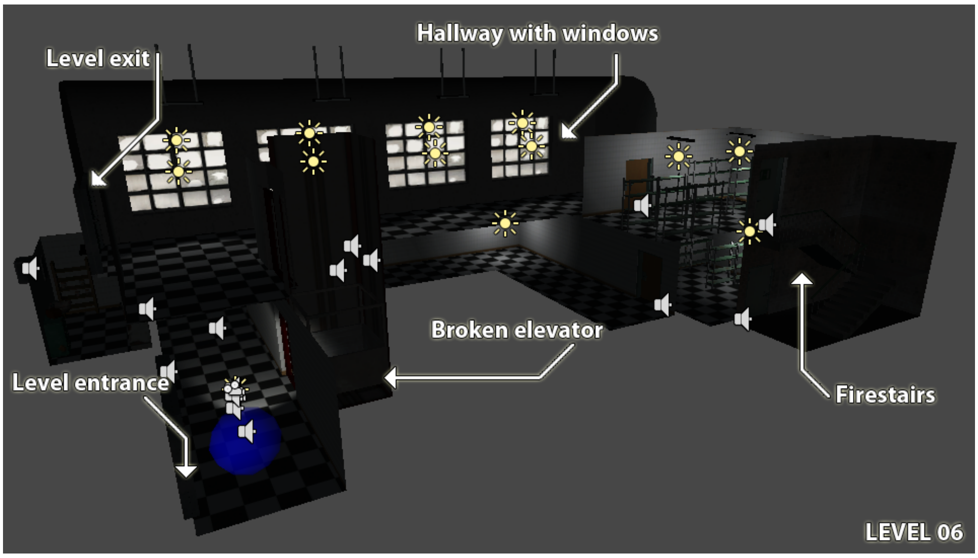




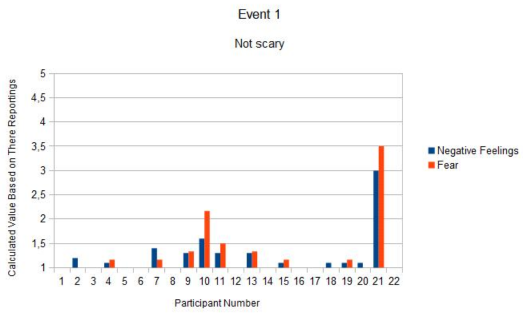
C Level 4-6 overviews

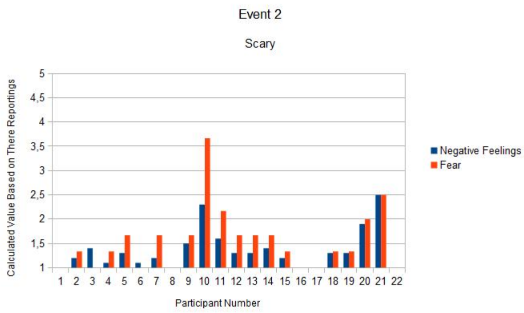


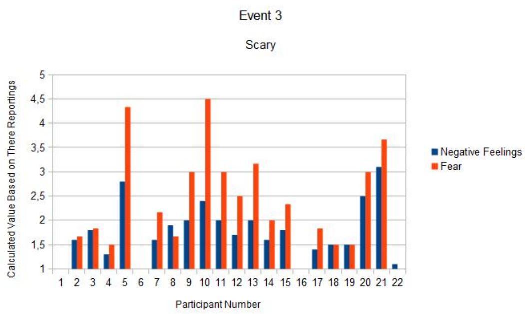


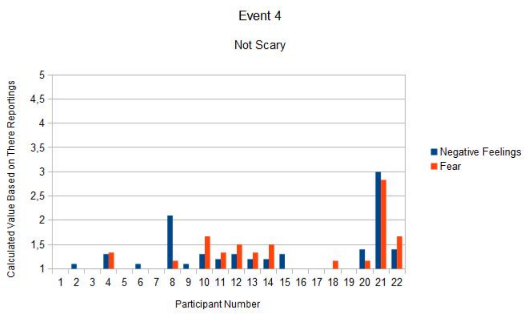


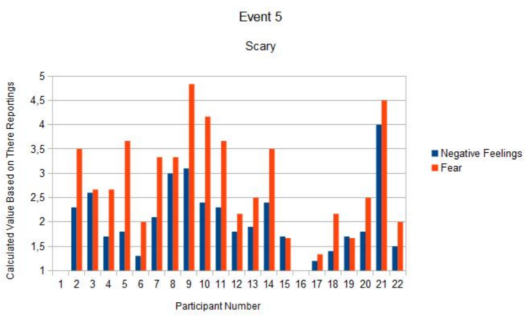
D Results for event 1 to 6 (Self-reporting)

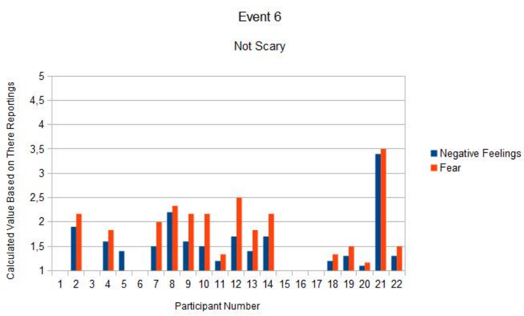




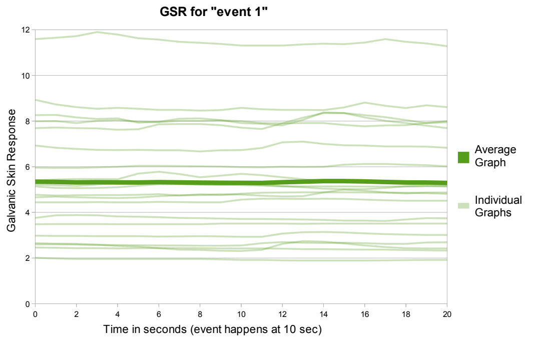


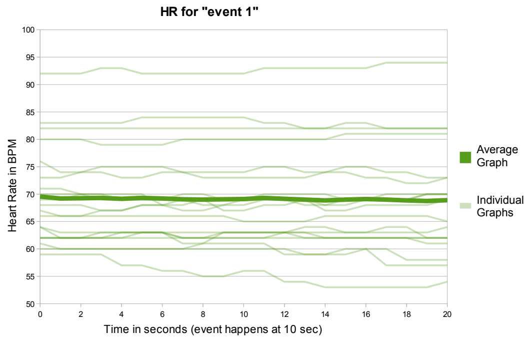


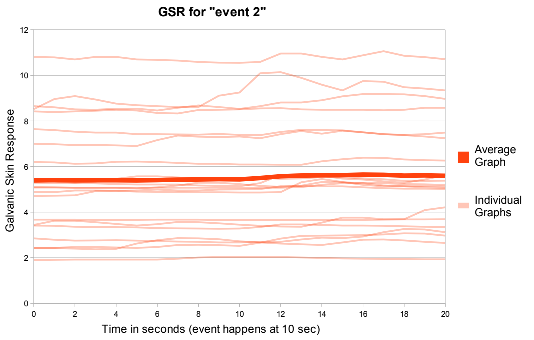


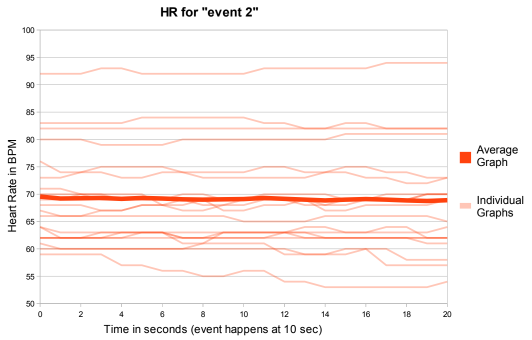


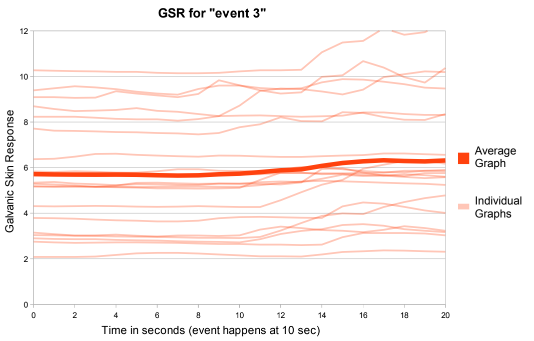
E Results for event 1 to 6 (Physiological data)

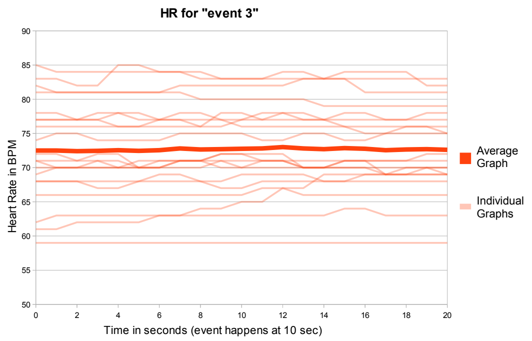


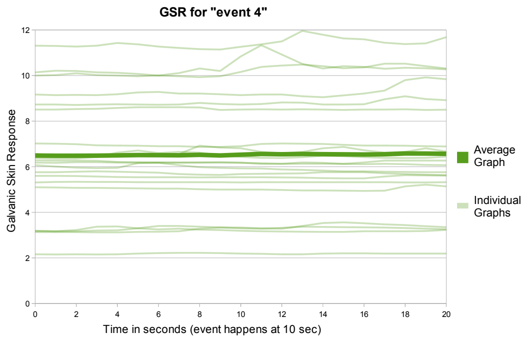


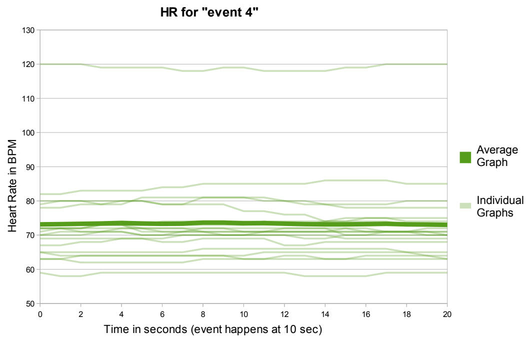


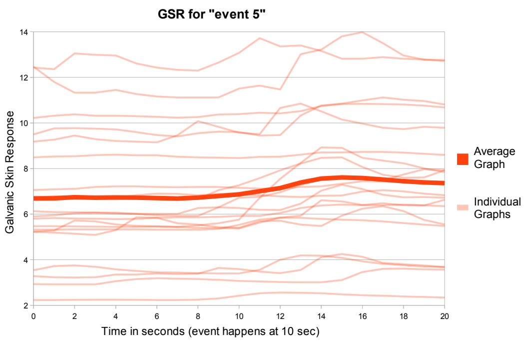


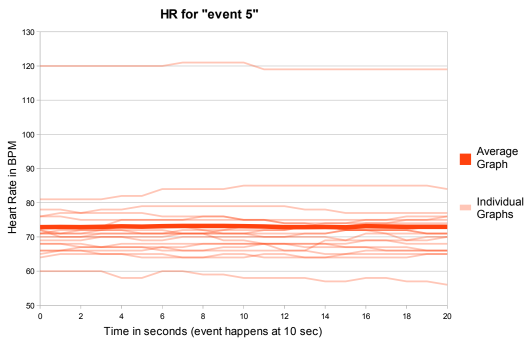


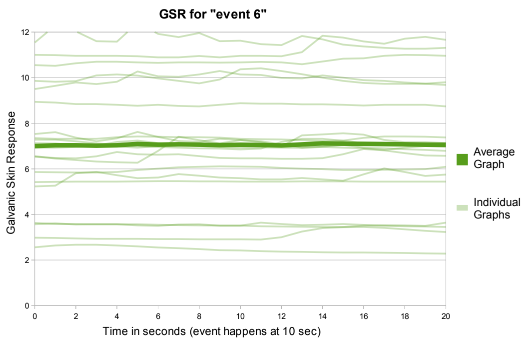


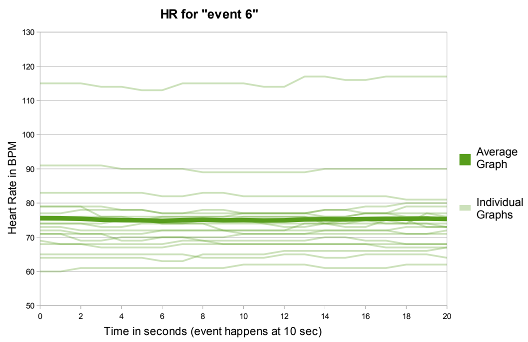

GSR for "event 4"
Galvanic Skin Response
Time in seconds (event happens at 10 sec)
Average Graph
Individual Graphs


HR for "event 4"
Heart Rate in BPM
Time in seconds (event happens at 10 sec)
Average Graph
Individual Graphs


GSR for "event 5"
Galvanic Skin Response
Time in seconds (event happens at 10 sec)
Average Graph
Individual Graphs


HR for "event 5"
Heart Rate in BPM
Time in seconds (event happens at 10 sec)
Average Graph
Individual Graphs


GSR for "event 6"
Galvanic Skin Response
Time in seconds (event happens at 10 sec)
Average Graph
Individual Graphs


HR for "event 6"
Heart Rate in BPM
Time in seconds (event happens at 10 sec)
Average Graph
Individual Graphs