The use of physiological measurements to assess user involvement in computer games

Du Nguyen, s082968

Kongens Lyngby 2013
DTU COMPUTE-2013

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253031
compute@compute.dtu.dk
http://www.compute.dtu.dk/ DTU COMPUTE-2013

Contents

1 Introduction **1**

2 Gaming Theory **3**

3 Emotion Mapping **7**

4 Physiological Measures **9**
4.1 EEG . . . 9
4.2 Other Physiological Measures . . . 22

5 Behavioral Measures **31**
5.1 GEQ . . . 31
5.2 SAM . . . 33

6 Conclusion **35**

A Index of Terms **39**

Bibliography **41**

Chapter 1

Introduction

The computer game industry is a massive industry. According to a report [ESA12] by the ESA (Entertainment Software Association) the total amount of money spent on computer and video games in 2011 was $24.75 billion in the US while it was $67 billion [Gau12] globally in 2012. It is clear that computer game industry is no longer a niche market but a large global market that will only continue to grow.

What is the secret behind the meteoric rise of the computer game industry? Often when speaking of entertainment and the lure of entertainment the term escapism is central. Merriam-Webster defines escapism as "habitual diversion of the mind to purely imaginative activity or entertainment as an escape from reality or routine" [MW13]. It can be argued that computer games provide a higher degree of escapism than movies or books because of their inherent nature of interactivity.

The quality of computer games are more or less defined of how fun they are to play. If a game is fun to play and provides that sense of escapism it often tends to score higher in reviews and sell more. The challenge in game development is how to create an experience that is fun and especially an experience that is fun for everyone.

The report proposes a system that monitors a player's physiological and mental state which can be used by games to dynamically adapt to the player and improve the experience for the player. This system can be used both in the development phase of the game and the "living room" phase. There are appealing reasons to use such a system for both the development phase and living room phase. For development this system can be used for play testing the game since this system returns objective metrics of the mental state of a player. For the living room phase the system can personalize the game experience and provide a smoother and individual game experience. As the living room system is more complicated it requires more thought in how such a system is designed. The rest of the report will deal with the issues of constructing a living room system such as: How can emotions be conceptualized? How can you measure fun and how can fun be created? How do you measure emotions and how do emotions relate to the feeling of fun?

Chapter 2

Gaming Theory

The ultimate goal of the current project is to increase game experience for computer game players. To achieve this, a theory of fun and game experience has to be defined and how these relate to emotions and a player's mental state. Knowing how game experience relate to emotions and how game experience is created will make it possible for the current project to alter the game experience for the player.

In an article Chen [Che07] describes how flow and games can be combined. Flow is a concept introduced by Mihaly Csikszentmihalyi which represents a feeling of complete and energized focus in an activity with a high level of enjoyment and fulfillment. Csikszentmihalyi identified eight major components of Flow:

- A challenging activity requiring skill
- A merging of action and awareness
- Clear goals
- Direct, immediate feedback
- Concentration on the task at hand
- A sense of control

- A loss of self-consciousness
- An altered sense of time

Flow can be induced by any kind of activity and Chen argues that most computer games include and leverage the eight components of Flow. Chen further argues that since all users have different skills and expectations an experience has to offer many different choices that adapts to the different user's Flow Zones. Chen concludes that to keep a user within the user's Flow Zone the game should offer choices that allow the user to enjoy Flow in their own way.

In Rules of Play [SZ04], Salen and Zimmerman also describe flow as one approach to describe pleasure in games. Like Chen, Salen and Zimmerman use the eight major components of Flow. They split it up in effects of flow and prerequisites for flow. The effects are:

- A merging of action and awareness
- Concentration on the task at hand
- A loss of self-consciousness
- An altered sense of time

And the prerequisites are:

- A challenging activity requiring skill
- Clear goals
- Direct, immediate feedback
- A sense of control

These four prerequisites are the ones that need to be maximized. Csikzentmihaly has furthermore distilled these four prerequisites into two dimensions: Challenge and ability. Figure 2.1 shows the two dimensions and the flow zone.

Csikzentmihaly shows that if a person has a high ability but the challenge is low then the person feels boredom. If a person is low in ability but the challenge is too great then that person will have a negative and intimidating experience leaving the player to feel anxiety. If the person is challenged to his/her ability then the person is in the flow zone.

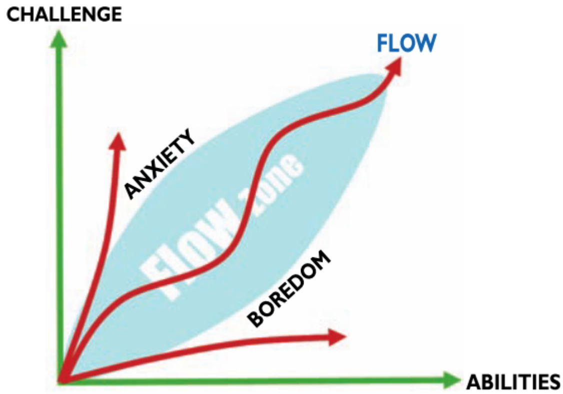

Figure 2.1: Flow zone in a challenge/ability chart.(From Flow in games [Che07])

Looking at the four prerequisites and their application in games Salen and Zimmerman discuss the importance of fulfilling these prerequisites. They emphasize the relationship between action and outcome. This relationship is presented by the clear goals and direct, immediate feedback from the prerequisites. The goals of a game must always be clear and the game must communicate where the goal is, how it might be achieved and whether the player is making progress toward it. The game must also make it clear for players what the outcome is to any certain action.

The two other prerequisites deal with challenge. Salen and Zimmerman argue that an anxiety state will result in the game feeling arbitrary. Therefore the challenge level must match the ability of the player. Salen and Zimmerman mention Dynamic Difficulty Adjustment or DDA as an approach to the management of challenge in games.

DDA is a technique that uses feedback loops to adjust the difficulty of play. An example is Crash Bandicoot where the player must maneuver through a series of jumping and dodging obstacles, overcome damaging hazards and reach objectives to finish a level. DDA is implemented in Crash Bandicoot by evaluating the number of times a player has died in a particular location in a level and adjusting the difficulty based on that criteria. If the player is having trouble the game automatically puts in a few more helpful objects or removes some enemies.

Hunicke implements a DDA system in the article "The Case for Dynamic Difficulty Adjustment in Games" [Hun05]. Hunicke describes several methods of implementing DDA in games. DDA can be manipulating the game economy, the narrative structure, the artificial intelligence or even the physical layout of maps or levels. Hunicke also notes that it is important when implementing DDA that

it must be unobtrusive. An example of badly implemented DDA is the Rubber Band effect which is often featured in racing games. The rubber band effect makes sure that the player does not get too far ahead of the computer which in racing games mean that no matter how far ahead a player is, the computer can always catch up by suddenly driving faster.
For the DDA implementation Hunicke is manipulating the game economy in an FPS. The game economy in an FPS is the amount of health of the player and enemies, amount of ammunition in the level, amount of health in the level and so on. By manipulating these currencies the game will become more or less challenging to the player which can be used to keep the player in the flow zone. The DDA is implemented by calculating the probability of player death. Based on this metric the game can adjust the difficulty by means of the game economy. The results show that the mean of player deaths decreased from 6.4 to 4.0 and that expert players report slightly elevated levels of enjoyment.
Hunicke concludes that even the small changes made in the game showed improvements in player performance without removing agency from the player.

For this project to work a clear method to improve game experience has to be implemented. Using the theory of flow by Csikzentmihaly a clear method for maximizing game experience is found by maximizing the time a player spends in the flow zone. Salen and Zimmerman identified four of Csikzentmihaly's eight major components of flow as being prerequisites for a flow experience. Csikzentmihaly has identified two dimensions on which the four prerequisites can be described: challenge and ability. The ability relies on the player's skill while challenge both relies on the player's ability but also on the level of challenge posed by the game. By having the game adjust the challenge according to the player ability the player can get into the flow zone. This is what Dynamic Difficulty Adjustment (or DDA) can do.
By having the difficulty adjusted dynamically, the challenge posed to the player can be changed dynamically. Hunicke presents a system for DDA in an FPS game. This system uses the game economy as a dynamic system in which the difficulty can be adjusted. The results showed better player performance and a slight improvement in enjoyment among expert players.
DDA has also been used several times in commercial games. Examples include: Half-Life 2 [Tol08], Max Payne [R* 12], Prey [Sie06], The Elder Scrolls IV: Oblivion [Hic11], Crash Bandicoot [SZ04], Jax and Daxter [SZ04], Left 4 Dead [Sal09] and more.
Hunicke's results and the use of DDA in commercial games show that DDA is a viable tactic for increasing game experience.

Chapter 3

Emotion Mapping

The current project requires a player's mental state to be identified before anything else can happen. For entertainment purposes the most important element is to tap into the emotions of a player. This necessitates the need for a terminology for emotions and the exploration of different emotion theories.
In a review by Hamann [Ham12] theories of emotion are described and the controversy between the theories are discussed.
There are two main theories of emotion. One where emotion can be conceptualized as discrete categories and one where emotion can be conceptualized as dimensions such as arousal and valence.
Discrete emotion theories usually propose a small number of emotions that are the most basic and universal emotions in humans. These emotions have specific characteristics and unique physiological and neural profiles. Emotions such as happiness, sadness, anger, disgust, fear and surprise are often mentioned as being the most basic emotions.
Dimensional emotion theories propose that emotions can be mapped onto a few dimensions where the most often dimensions mentioned are arousal and valence. Arousal is a measure of how alert a person is and ranges from calm to excited. Valence is a measure for how pleasurable an event is to a person and ranges from highly positive to highly negative.
Hamann describes studies which study the discrete emotion theory. The results point towards limited evidence for consistent associations between brain regions and basic emotions. Rather the results indicated that one brain area could be

activated by different emotions instead of one brain area per emotion.
Hamann writes that meta-analyses of the dimensional emotion theory of arousal and valence show that representation of arousal and valence is quite complex. The studies show that arousal and valence both involve multiple brain regions and that these dimensions are not independent.
Hamann states that a one-to-one mapping of emotions to brain regions is not possible and that a network model where each emotion is a network of multiple brain regions is more correct.
The Hamann report suggests that a network view of emotions is a future direction for studies in emotion representation. Currently the choice for the current project is between basic emotions or a dimensional emotion representation since it is too early to be using a network view of emotions. The network view furthermore requires fMRI which requires a person to lie still inside a large machine and therefore limits it to clinical studies. Since a network view of emotions requires a high spatial resolution, it suggests that EEG would not useful since spatial resolution is one of the weaknesses of EEG. This leads to the conclusion that the valence/arousal model is clearly a better choice. This is due to the simplicity of having two parameters, valence and arousal, that can be tweaked. Furthermore it is widely accepted that arousal correlates with skin conductance response and valence with facial muscle activity. This simplifies the process of reading the emotion of a person and combined with the fact that the signals from these sensors are stronger than EEG signals this makes the valence/arousal model more suited for the current project.

Chapter 4

Physiological Measures

Since the current project is about measuring a player's mental state and emotions a clear approach to measuring the emotions of a player has to be found. Two approaches are described below. The first is measuring electrical activity on the scalp also known as EEG. The second is physiological measures that does not include EEG such as skin conductance, heart rate and temperature. It is important to correctly assess the emotions of the player and for this a clear and precise method of measuring emotion has to be found. In the following sections both approaches are described and results of the approaches are presented and compared.

4.1 EEG

Electroencephalograhy or EEG is the recording of electric activity on the scalp. The electrical activity corresponds to neuronal activity in the brain. The use of EEG for detecting emotions in games has not been a large research topic. For this report four studies are presented. A study by Chanel [CKGP06] and a study by Bos [Bos06] both look at recognizing emotions using images as stimuli. Both of these studies use images from the International Affective Picture System [Uni11] or IAPS which is an image database compiled by Bradley and Lang where all images are tagged with valence and arousal. The two other studies uses games as stimuli – Chanel 08 [CRBP08] and Reuderink 13 [RMP13]. Interestingly all studies uses the valence/arousal model of emotion with the Reuderink study using an additional dominance dimension.

The Chanel 06 [CKGP06] study details an experiment where 4 participants were asked to look at images from the IAPS database. After each image the participants had to self-assess their valence and arousal using the self-assessment manikin [BL94] or SAM which is a non-verbal pictorial technique to measure valence, arousal and dominance. The participants wore an EEG device, a galvanic skin response (GSR) sensor for skin conductance, plethysmograph for blood pressure measurements, respiration belt for abdominal and thoracic movements and a temperature sensor. The participants were asked to look at 100 images each – 50 images that were tagged as low arousal in the IAPS database and 50 images that were tagged as high arousal.
The data was processed using both a Bayes classifier and a classifier based on Fischer Discriminant Analysis (FDA). The classifiers were trained for each participant with leave one out cross validation where one pair of classifiers (Bayes and FDA) used two ground-truth classes and another pair used three ground-truth classes. The ground-truth classes were constructed from the SAM which contained 5 steps of arousal from calm to exciting.
The results show a mean over all subjects as 60% for the Bayes classifier using only EEG features and two ground-truth classes while it is 55% for the FDA.
Using three ground-truth classes the mean for the Bayes classifier using only EEG features is 45% while for FDA it is 40%.
Figure 4.1 shows a plot of the reults.
Chanel concludes that EEG signal seem to perform better than physiological signals and that the concatenated signals (both EEG and physiological) with FDA shows the best performance. Another conclusion from the study is that using the self-assessments as ground-truth classes were better than using the IAPS ratings.

The results from the study show unconvincing results. For all of the classification problems the mean over all participants are barely above chance level and for some participants the mean is below chance level.
Apart from unconvincing results the results also show large variance with participants below chance level and other participants well above chance level.
Another problem is the fact that the concatenated signals are not equal to or better than the best non-concatenated signal. This is seen in Figure 4.1 in the top left plot. Here the EEG signal is better than the concatenated signal which is odd since the concatenated signal also contains EEG signals and could therefore choose to disregard the physiological signals to get the same performance of the EEG signals.
The study also builds a classifier for each participant which makes it less useful

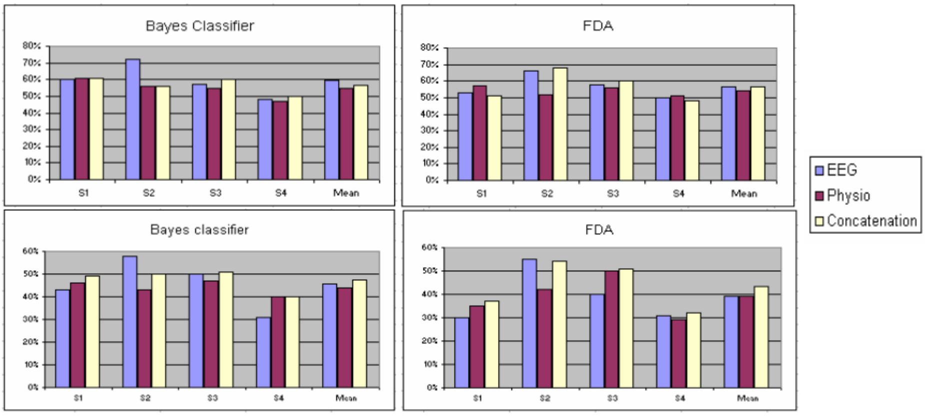

Figure 4.1: The plots in the top row are the results for classifying into two classes. Bottom row plots are for the result for classifying into three classes. Left column is the plot for Bayes and right column is for FDA. Along the axes are subjects and mean on the x-axis and how correct the classifier is in percentages along the y-axis. The blue bar is classification from EEG signals, the red bar is classification from physiological signals without EEG signals and the beige bar is the fusion of EEG signals and physiological signals.

for practical applications.

Reuderink [RMP13] describes an experiment where the goal is to confirm or reject hypotheses about correlations between EEG signals and arousal, valence and dominance. The experiment was conducted with 12 participants who played a rigged pacman game. The participants played 30 minutes of the pacman game in 2-minute blocks wherein one third of the 2-minute blocks would ignore 15% of key presses and therefore induce frustration. After each block the participants were asked to fill out a SAM.
The participants were wearing an EEG advice, EOG for measuring eye position and an EMG device for measuring muscular activity. Other sensors were also worn but the data was not used. The EOG and EMG were used for finding ocular and muscular artifacts.
The results of the SAM which can be seen in Figure 4.2 indicate that frustration leads to low valence and low dominance.
The study also looks at the relation between time, experimental condition, valence, arousal and dominance which is gathered from the SAM. A table of correlations can be found in Figure 4.3. The study shows that there are no correlations with time and anything else which suggests that the emotional ratings do not drift over time. There is also a correlation between valence and dominance which according to the study should not be since valence and dominance are supposed to be orthogonal.
The focus of the study is however to study several hypotheses regarding to measuring valence, arousal and dominance. Most notably Reuderink studies the hypothesis that left-right brain alpha asymmetry as a measure of valence. That is the asymmetry in the alpha frequency band between the left and right brain hemispheres is related to valence. A full list of hypotheses can be seen in Figure 4.4.
The results of testing the alpha asymmetry hypothesis can be seen in Figure 4.5 Reuderink concludes that they found correlates for valence and arousal in theta, delta and alpha bands. Reuderink also confirms that alpha asymmetry can be used as a measure for valence.

The results in both Figure 4.2 and 4.3 show some interesting observations in which gameplay dominance seems to correlate with valence while valence and arousal does not correlate.
This study shows that the dominance dimension is not needed for the current project because the dominance and valence dimensions correlate. This is even more relevant when taking into account that the environment of the study was a computer game because it shows that the valence/arousal model is good enough for computer games and that the dominance dimension is unneeded since it correlates with the valence dimension.
The correlates found in the study are not very strong. The correlates can be found in Figure 4.6. Here the correlations typically range from 0.2 to -0.2 and

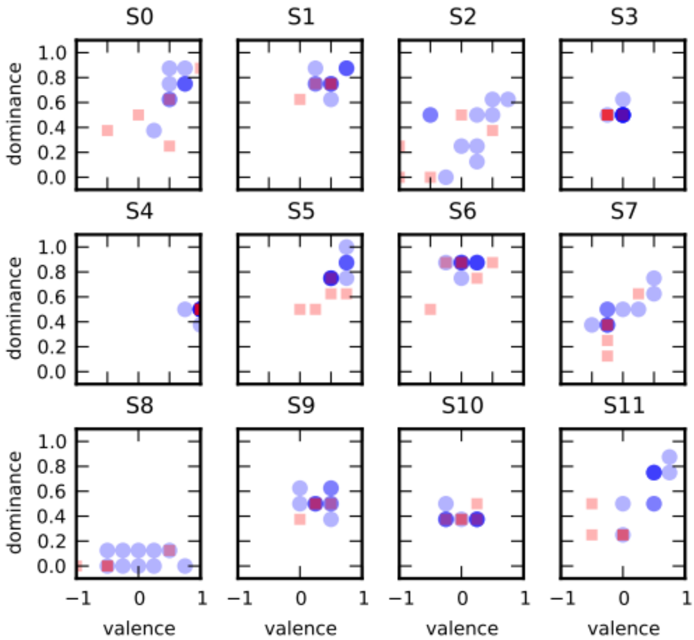

Figure 4.2: Valence and dominance ratings plotted for each participant. The normal play sessions are denoted by blue circles, frustration sessions are denoted by red squares.

	Time	Cond.	Val.	Ar.	Dom.
Time	1.00	0.01	−0.04	0.07	0.01
Condition	0.01	1.00	−0.32	0.08	−0.32
Valence	−0.04	−0.32	1.00	0.10	0.43
Arousal	0.07	0.08	0.10	1.00	−0.15
Dominance	0.01	−0.32	0.43	−0.15	1.00

Figure 4.3: Table of correlations.

Dimension	Delta	Theta	Alpha	Beta	Gamma
Valence ↑	$H_{v\delta}$: fron.-med. ↑	$H_{v\theta1}$: l-hemi. ↑ $H_{v\theta2}$: r-hemi. ↓ $H_{v\theta3}$: fron.-med. ↑	$H_{v\alpha1}$ l-hemi. ↓ $H_{v\alpha2}$ r-hemi. ↑	-	$H_{v\gamma1}$: l-temp. ↓ $H_{v\gamma2}$: r-temp. ↑
Arousal ↑	$H_{a\delta}$: posterior ↑	$H_{a\theta}$: posterior ↑	$H_{a\alpha1}$: frontal ↑ $H_{a\alpha2}$: global ↓	$H_{a\beta}$: parietal ↑	$H_{a\gamma}$: gamma ↑
Dominance ↑	-	-	-	-	-

Figure 4.4: Table of hypotheses for valence, arousal and dominance. Note that no hypotheses for dominance are listed.

	Fp1-Fp2	AF3-AF4	F3-F4	FC1-FC2	C3-C4	F7-F8	P3-P4
Condition	$-0.15*$	-0.04	-0.07	-0.00	-0.09	-0.05	0.05
Valence	0.19	0.00	0.12	$0.16*$	0.11	-0.05	0.06
Arousal	0.00	-0.13	-0.12	$-0.16*$	0.08	-0.04	0.06
Dominance	0.07	0.01	$0.18*$	$0.18**$	-0.06	0.03	-0.01
PC0	$0.19*$	0.04	0.08	0.05	0.11	0.02	-0.01
PC1	0.03	0.04	0.03	-0.07	0.03	0.10	-0.12
PC2	-0.01	0.13	$0.17*$	$0.19**$	-0.13	0.05	-0.04
PC3	-0.05	-0.10	-0.08	-0.14	-0.04	-0.00	0.02

Figure 4.5: Table of correlations for alpha asymmetry for different sensor pairs. PC0-3 indicate principal components where PC0 more or less corresponds to positive emotions, PC1 corresponds to negative emotions, PC2 to relaxed and dominance, PC3 to aroused and dominance. One asterisk indicates $p<0.05$, two indicates $p<0.01$

this is and seeing as the EEG can show both positive and negative correlations right next to each other (e.g. valence at 3 Hz) it shows that EEG does not provide good enough measurements and that the spatial resolution on EEG devices cannot provide enough detail as mentioned in the Hamann [Ham12] study.

Reuderink confirms that alpha asymmetry can be used as a measure for valence although looking at the valence row in Figure 4.6 it looks to be a weak correlation. Figure 4.5 also shows weak correlates. Furthermore Reuderink concedes that for most of the hypotheses no significant correlations were found and this greatly reduces the usage of EEG as an emotion recognizer.

The Bos 06 [Bos06] study is similar to the Chanel 06 study. The experiment in this study was one where 5 participants had to look at 12 images from the IAPS database, listen to 12 sounds from the International Affective Digital Sounds (IADS) database which are sounds rated for valence and arousal, and lastly look at and listen to 12 audiovisual stimuli which are IAPS images and IADS sounds combined.

Each participant was wearing an EEG device and the resulting data was used for classification with FDA. The classification is for two classes – high and low arousal or valence and a classifier was trained for each participant using 3-fold cross validation.

The results show the average classification rate for arousal to be around 60-70% with the best classification rate around 90%. For valence the results are 60-70% average classification rate and around 90% for the best classification rate.

The results for arousal classification are seen in Figure 4.7 and for valence in Figure 4.8.

Bos concludes that their good results would not be equally impressive if the classified arousal and valence has to be used to classify emotions as this could

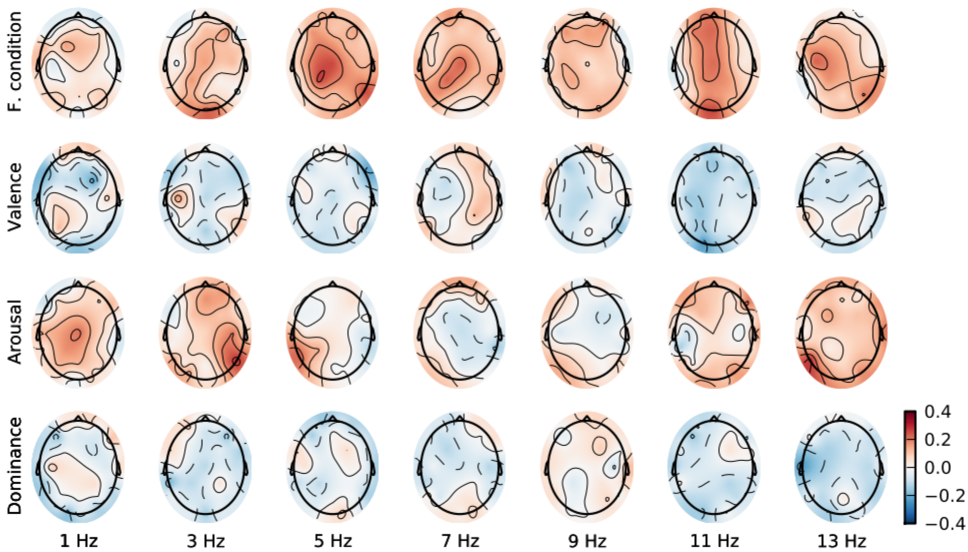

Figure 4.6: Mean subject correlations where the rows indicate frustration, valence, arousal and dominance while the columns is the frequencies. Each circle represent a spatial view of a head where colors indicate correlations. For example does arousal at a frequency of 3 Hz show a correlation at about 0.3 at the right back of the head.

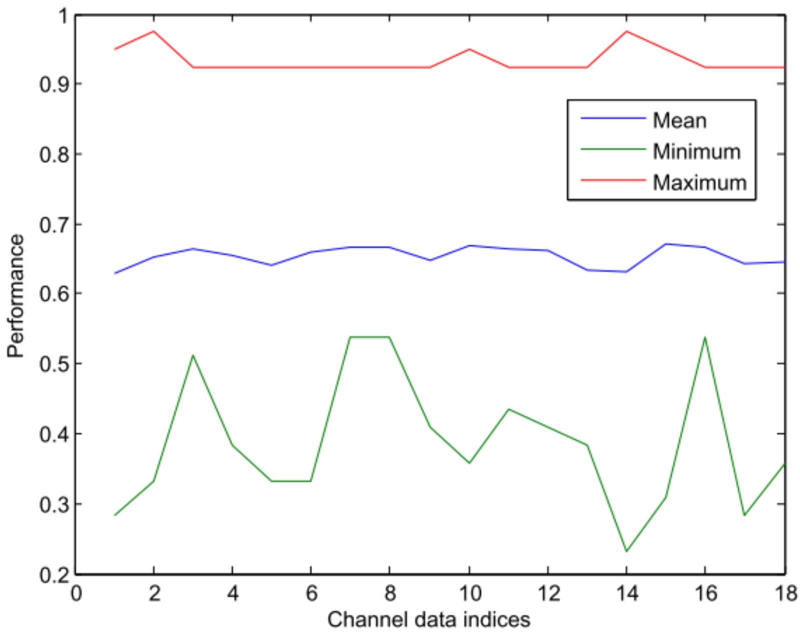

Figure 4.7: Performance of each channel feature for arousal classification. The performance is percentages correct for classification in two classes – high arousal and low arousal. The channel data indices refer to features which are combinations of frequency bands over the EEG channels. That means channel 1 is a feature which is a certain frequency band measured on a specific EEG channel.

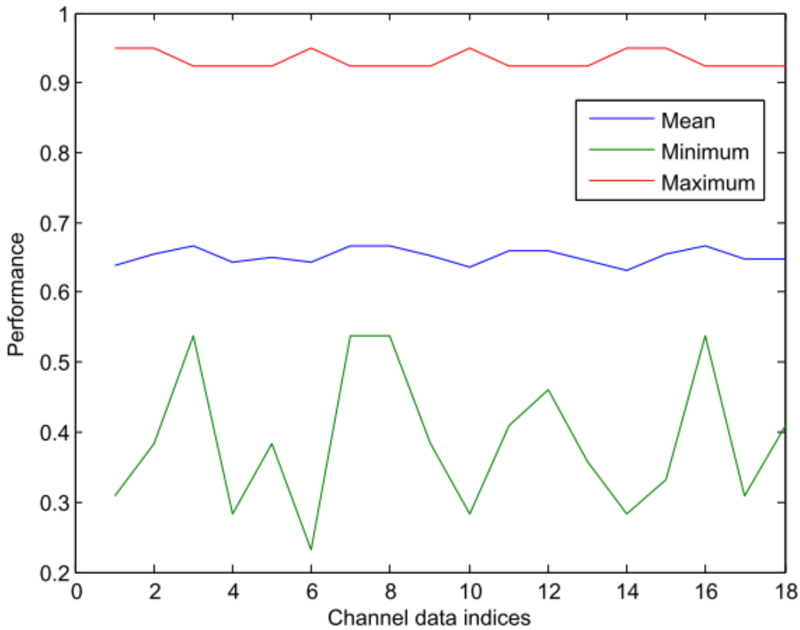

Figure 4.8: Performance of each channel feature for valence classification. The performance is percentages correct for classification in two classes – high arousal and low arousal. The channel data indices refer to features which are combinations of frequency bands over the EEG channels. That means channel 1 is a feature which is a certain frequency band measured on a specific EEG channel.

introduce more errors.

The study by Bos is on close inspection not good enough. One problem is that the channel data indices are not stated anywhere although the conclusion does provide the frequency band and location of the electrodes for the best performing channel.
The main problem is however that the worst performing classifier is just above chance level for some few frequency band and location of electrodes combinations. This is a worry since the mean is around 60-70% and the maximum is at 90%. That indicates large variance in the data or at least some outliers.
Another smaller problem is the fact that the ground-truth classes for the classifiers are based on the IAPS/IADS ratings which is a problem because the images and sounds will not have the same effect on people. This is something that Chanel found out in the Chanel 06 study where he decided to bypass the IAPS ratings and solely use the SAMs.
As seen in the Chanel 06 study a classifier was trained for each participant which again makes the usage of EEG or at least the method from this report less practical.

Another study from Chanel[CRBP08] describes an experiment where emotions are measured using physiological measures and a computer game as stimuli.
Using the psychology of Flow Chanel sets out to measure three states: anxiety, flow and boredom. Since direct measurement of these three states cannot be done Chanel has mapped the three states to the valence/arousal theory of emotion. Anxiety is therefore negative valence, high arousal, flow as positive valence and high arousal and boredom as negative valence and low arousal.
The experiment was conducted with 20 participants where each participant wore an EEG sensor, plethysmograph, respiration belt, GSR sensor and a temperature sensor. The game was Tetris where each participant would play for a while to determine their skill level. The skill level would equal a medium difficulty and easy and hard difficulty was calculated from the medium difficulty. After every 5 minutes of play the participants filled out a SAM and a 30-question questionnaire. The participants would play two 5 minute sessions of each difficulty level.
The data was classified into three classes of bored, flow and anxiety using support vector machines and using signals from all sensors except EEG. The classifier for each participant was trained using features from the other participants.

The results of the SAMs show no surprises in that the medium difficulty shows the highest valence and medium level of arousal. The arousal increases as the difficulty increases. These results can be seen in Figure 4.9.

The results of the classification are presented in Figure 4.10. The overall classification rate is 53.33% for 3 classes though there are quite significant differences

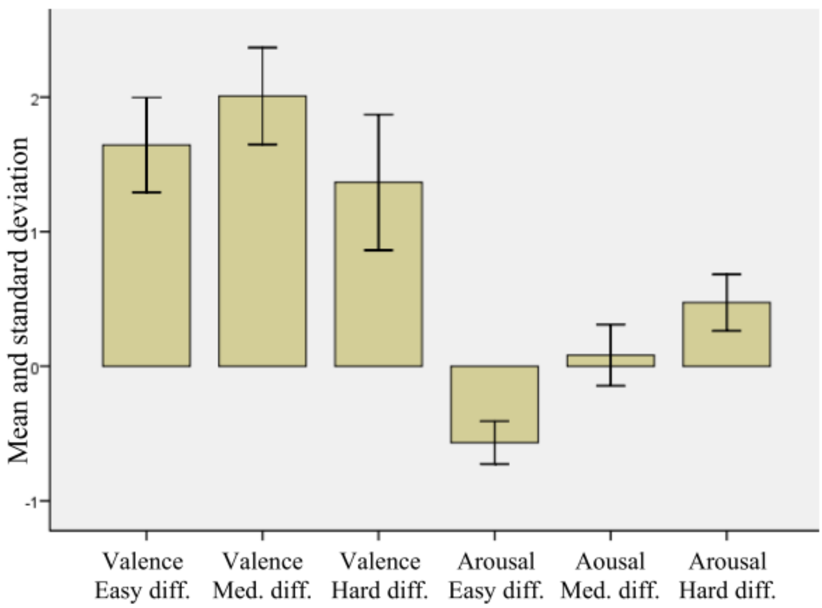

Figure 4.9: Mean and standard deviation for valence and arousal in the three difficulties. Valence and arousal values are from questionnaires.

Classified / True	Easy (Boredom)	Medium (Engagement)	Hard (Anxiety)
Easy (Bored.)	72.5%	20.0%	7.5%
Medium (Eng.)	37.5%	20.0%	42.5%
Hard (Anxiety)	29.0%	2.6%	68.4%

Figure 4.10: Confusion matrix for the classification of the three classes.

between each category. For boredom the classification rate is 72.5%, flow is 20% and anxiety is 68.4%.

Chanel concludes that the classification rates for two of the emotions, boredom and anxiety, are good. Chanel also concludes that the classification of flow is unnecessary since it tended to be classified as boredom or anxiety instead and proposes only measuring for boredom. Chanel argues that while boredom is unwanted in games, anxiety is not entirely unwanted as a greater challenge than the player's ability can stimulate the player to learn and perform better at the game.

The study shows good results although EEG was not used. The classification rates for boredom and anxiety are great although there is a worry with the classification rate for the flow state. The classification rate for the flow state is 20% which means that the classifier is more likely to classify the flow state as either boredom at 37.5% or anxiety at 42.5%. This is a worry but as Chanel argues it might be a better solution to only detect boredom and perhaps extreme frustration or anxiety.
Furthermore classification was done using other participants as training set which means that the classifier and results are not user specific. This is very useful for the current project since it is cumbersome to train a classifier each time a user wants to use the system.

The two studies, Chanel 06 and Bos 06, do not show good results which brings the usability of EEGs for emotion recognition into question. The Chanel 08 study shows good results but the classification only uses physiological signals other than EEG and as such it also suggests that EEG is not suited for emotion recognition.
The studies that do use EEG for emotion recognition do not provide good results. The Chanel 06 study shows unconvincing results. For classification in two classes using Bayes classifier the mean of all participants using EEG is 60%. That is 10 percentage point better than chance level. For classification

Study	Class. classes	Chance level	Findings	PP diff.
Chanel 06	2	50%	60%	10
Chanel 06	3	33%	45%	12
Bos 06 (arousal)	2	50%	65%	15
Bos 06 (valence)	2	50%	65%	15

Table 4.1: Comparison of EEG results. PP is an abbrev. for percentage points

into three classes using Bayes classifier the result is 45% which is 12 percentage points better than chance level. A comparison is shown in Table 4.1.
The Reuderink 13 study is not about classification so it cannot be easily compared with the other studies. The study instead finds weak correlates between specific brain areas, frequency bands and valence, arousal and dominance which does not indicate that EEG is a good enough solution. Furthermore Reuderink concedes that for many hypotheses few could be confirmed.
The Bos 06 study shows some confusing results. In classification with two classes using binary linear FDA the classification rate for both valence and arousal is around 65%. It is therefore 15 percentage points better than chance level. The downside is however that there seems to be large variance in the data since the minimum performance is around 50% and in most cases below 50% while the maximum performance is consistently around 90%.

For both Chanel 06 and Bos 06 studies the percentage points difference between chance level and findings are 10-15 percentage points. That is a quite modest result and on closer inspection it turns out that the two studies show large variance. For the Bos 06 study minimum performance is around 50% or a chance level and the maximum performance is around 90%. The Chanel 06 study has participants that perform worse than chance level and some who perform much better than the average. This is a major worry that not only are the results modest, the variance seems to indicate that performance could be person dependent.
One issue with EEG for emotion recognition is that there is no consensus on which frequency bands and brain areas that represent specific emotions. This would greatly complicate the current project. Furthermore as Hamann states, the usage of EEG for emotion recognition is not recommended since results from other studies show that emotions might be represented by a network of brain areas and not just a single brain area. This would require some spatial view of the brain which is one of the weaknesses of EEG and together with the results from the Chanel 06,Bos 06 and Reuderink 13 studies, using EEG is not recommended for the current project.

4.2 Other Physiological Measures

The human body is a massive signal generator. These signals can be measured with various instruments. The ones presented are candidates for emotion recognition in game applications.
GSR or Galvanic Skin Response which is also known as electrodermal activity or skin conductance level is a technique where a device will measure the conductivity of the skin through the sweat level of the skin. This is useful because when people experience physical arousal sweat is produced which leads to better conductivity.
HR or Heart Rate is measured through a heart rate monitor but can also be measured using many different methods.
BVP or Blood Volume Pulse is a method for measuring blood flow through skin capillary beds in the finger. A BVP sensor is put on a participant's finger and gives a measure on anxiety. This is due to the "cold feet" phenomenon where blood drains from the extremities a person during periods of emotional duress. The BVP sensor can also be used for HR measurements.
Another physiological signal is muscle activity. This is measured with EMG or Electromyography. EMG detects surface voltages when a muscle is contracted [MIC06]. In emotion recognition EMG are often used on facial muscles to detect smiles and frowns.
In this section four studies are presented. One by Drachen [DN10] is about finding correlates between physiological signals and game experience. The three others, Lisetti 04 [LN04], Kim 04 [KBK04] and Scheirer 02 [SFKP02], are about emotion recognition. The use of stimuli varies, Drachen and Scheirer use computer games, Lisetti uses movie clips and Kim uses audio, visual and cognitive stimuli.

The Drachen study [DN10] describes an experiment wherein 16 participants were playing 3 computer games (all in the First-Person Shooter genre) while wearing a HR monitor and a GSR device. Every 5 minutes the participants were asked to complete a questionnaire called iGEQ or ingame Game Experience Questionnaire. The iGEQ is a self-report scale for player experience and contains seven dimensions of player experience: Immersion, Flow, Competence, Tension, Challenge, Negative affect and Positive affect.
The data was filtered and the correlations were calculated with Pearson's correlation coefficient.
The results shown in Figure 4.11 show that there is a correlation between low heart rate and positive affect, flow, low challenge, immersion and feelings of competence. There is a correlation between high heart rate and tension and negative affect.
For GSR (EDA in the figure) there is a correlation between low skin conductivity and immersion, flow and positive affect while there is a correlation between

Physiological measures	Competence	Immersion	Flow	Tension	Challenge	Negative affect	Positive affect
HR	-0.36	-0.43	-0.25	0.37	-0.31	0.24	-0.42
EDA	-0.08	-0.23	-0.24	0.02	-0.18	0.38	-0.20

Figure 4.11: Pearson correlation coefficient for the seven dimensions of iGEQ and HR/GSR.

Classifier	Sadness	Anger	Surprise	Fear	Frustration	Amusement
KNN	70.4%	70.8%	73.9%	80.9%	78.3%	69.6%
DFA	77.8%	70.8%	69.6%	80.9%	72.7%	78.3%
MBP	88.9%	91.7%	73.9%	85.6%	77.3%	87.0%

Table 4.2: Comparison for classification rates of each classifier.

high skin conductivity and negative affect.

Drachen concludes that physiological measures correlate with game experience although it is dependent on context and approach.

The results indicate that game experience does correlate somewhat with physiological signals. One worry is that the correlations are not strong although it might not be a problem for the current project but this may not prove a problem since the correlations are for the seven dimensions of game experience. Positive and negative affect does however show the strongest correlation which is good since positive and negative affect could relate a lot with valence.

The Lisetti study [LN04] attempts to recognize emotion from physiological signals using movie clips as stimuli. The experiment had 29 participants watch movie clips while having their HR, GSR and temperature measured. Another study had preceded this where the movie clips were chosen. 5 movie clips were chosen from that study to represent sadness, anger, amusement, fear and surprise and the participants were also asked to solve a difficult mathematical problem without aids which was to represent frustration. After each clip/task the participants were asked to fill out a questionnaire about whether the participant felt the intended emotion, rate the intensity of that emotion and whether other more intense emotions were felt.
The data resulted in 12 features and for each of the six classes three classifiers were trained. The three classifiers were k-nearest neighbor algorithm (KNN), discriminant function analysis (DFA) and the Marquardt backpropagation algorithm (MBP). The KNN was tested with leave one out cross validation.
The results are shown in Table 4.2 and the results for the best classifier, MBP, are shown in Figure 4.12.

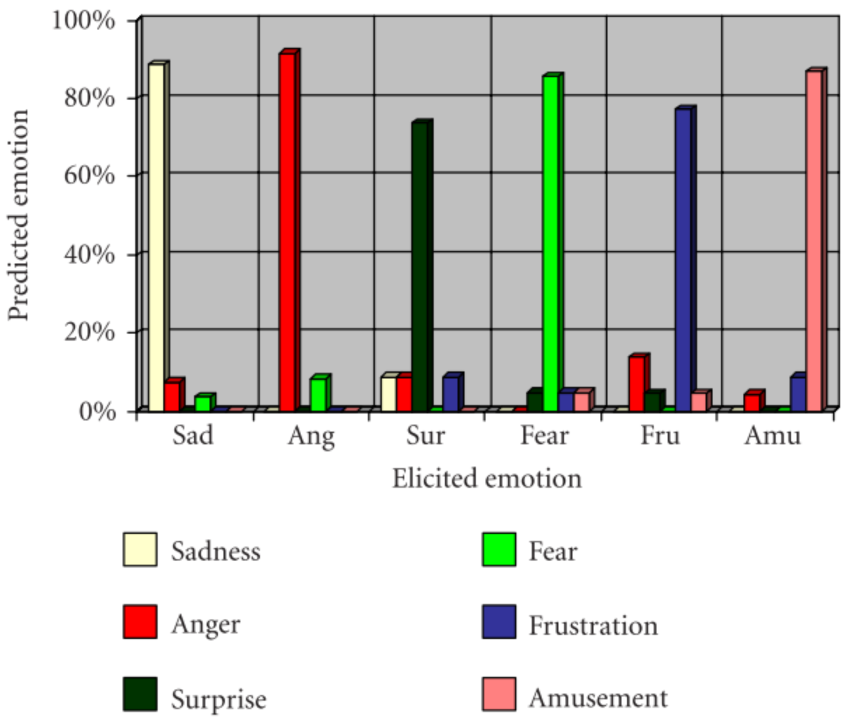

Figure 4.12: The results of classification using the MBP algorithm. For each emotion on the x-axis there is a bar for each emotion that has been classified. The height of the bar corresponds to the percentage that the emotion has been classified.

The results showed that the DFA algorithm was better than KNN for sadness, frustration and amusement, KNN was better for surprise and MBP was best in all classes except surprise and frustration.
The study concludes that the three algorithms can categorize emotions with 72.3%, 75.0% and 84.1% accuracy for KNN, DFA and MBP, respectively.

The study present some great results since the average classification rates are 72.3%, 75.0% and 84.1% for a six class problem.
There are however a problem that could have had an effect on the classification rates. The problem is that Lisetti only use one clip to represent each emotion. This could lead to the classifier to classify according to which clip was shown and not an emotion. Lisetti asked the participants whether or not they thought the clip evoked the intended emotion. For some of the clips the agreement rate are around 50-60%. This means that a specific clip only invoked the intended emotion in half of the participants or they associated another emotion with the movie clip. In connection the classifier could have classified according to the physiological footprint of the clips instead of classification of the emotion felt. As such this study cannot definitely show that GSR, HR and temperature are good choices for emotion recognition.
Finally the KNN classifier was used with leave one out cross validation which results in a classifier that is not user specific.

The study Kim 04 [KBK04] describes an experiment where participants were subjected to auditory, visual and cognitive stimuli in an attempt to recognize emotion. The participants were 175 children in the age of 7-8 years and the experiment was conducted in two occasions with 125 children in the first and 50 children in the second. The children had their GSR, HR and temperature measured.
The first experiment was to recognize sadness, anger and stress while the second was to recognize the three emotions mentioned and the emotion surprise. Apart from measuring an additional emotion the second experiment was also conducted with instructions for the children to be as still as possible.
The data was classified with support vector machines and the result of the second experiment is shown in Figure 4.13.
The results for the first experiment returned a classification rate of 55.2% for three classes. The second experiment returned a classification rate of 78.4% for three classes and 61.8% for four classes.

Kim concludes that a novel emotion recognition system has been developed based on the physiological signals of HR, GSR and temperature since the classification rate is significantly higher than chance level.

The results presented are great with classification rates of 78.4% for three classes and 61.8% for four classes. This is from the second experiment and the difference

	Recognition result (three emotional statuses)		
Original status	Sadness	Anger	Stress
Sadness	10	4	3
Anger	0	16	1
Stress	2	1	14

	Recognition result (four emotional statuses)			
Original status	Sadness	Anger	Stress	Surprise
Sadness	11	2	3	1
Anger	0	13	4	0
Stress	1	5	10	1
Surprise	3	4	2	8

Figure 4.13: Recognition results for the second experiment. The numbers indicate number of participants.

between the first and second experiment can be a worry. For the current project the system is intended to be used in a living room scenario and if the system places restrictions then it might not be widespread. For the second experiment the children were asked to sit as still as possible and the results improved from 55.2% to 78.4%. In reality it might not pose a problem but it is something has to be considered before the current project can be released.
An issue is that the classification was based on a training set consisting of a third of the entire sample. The classification rate was then calculated on the rest of the sample. What would have been interesting is if Kim ran cross-validation on the data set. This is however a minor problem since Kim did use one third of the children as training and the rest as testing. This ensures that the classifier is not user specific.
Questions could also arise on the use of children in the study. Would the results change if adults were used instead?
Although results are not as great as the Lisetti 04 study it does point toward the results in the Lisetti 04 as being representative of emotion recognition.

Scheirer 02 [SFKP02] is a study that attempts to classify whether a participant is frustrated or not. In the experiment conducted by Scheirer 36 participants were to play a computer game where the best participant would receive a cash prize. Frustration was induced by rigging the mouse to ignore clicks and the participants wore a GSR sensor and a blood volume pulse sensor.
The data was classified with Hidden Markov Models (HMM) using data from 24 participants. The ground-truth classes were frustration and non-frustration. The frustration classes consisted of the 10 seconds after a frustration event and

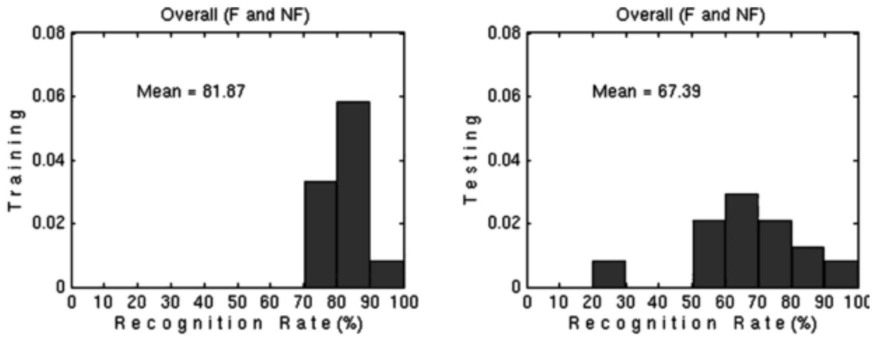

Figure 4.14: Overall classification rate for frustration and no frustration. Results are from 24 participants and therefore from 24 HMM structures. The plot on the left are classification rates for the training and the right are for the testing.

the rest of the samples were labelled as non-frustration. A HMM was trained for each participant using the data from the other participants.
The results which can be seen in Figure 4.14 show an overall classification rate of 81.87% for the training set and 67.39% in the testing set. The classification rate for frustration is 53.26% and for no frustration the classification rate is 72.44% which can be seen in Figure 4.15. Scheirer concludes that results might have been influenced by the labelling of ground-truth classes. The strategy of labelling the following 10 seconds after a frustration event is perhaps not the best as evidenced in Figure 4.15, top right plot where some show a recognition rate well below chance level. Scheirer however also concludes that for 21 out of 24 participants the classifier could classify frustration and non-frustration with a classification rate above chance level.

The results are very modest. For recognizing frustration it only performs slightly above chance level with 53.28%. Overall classification rate is 67.39%. The study mentions the ground-truth classes as a problem and this can perhaps be seen in Figure 4.15 where the classifier completely fails to recognize frustration for some participants with classification rates below 30% and well below chance level. Although the classification rate is 67.39% that results might improve with better ground-truth class labelling.

Overall the four studies using physiological other than EEG present some good results.
The study by Drachen seeks to find correlates between physiological signals and player experience in computer games. They succeed somewhat, finding moderate correlates between HR, GSR and player experience. The results are

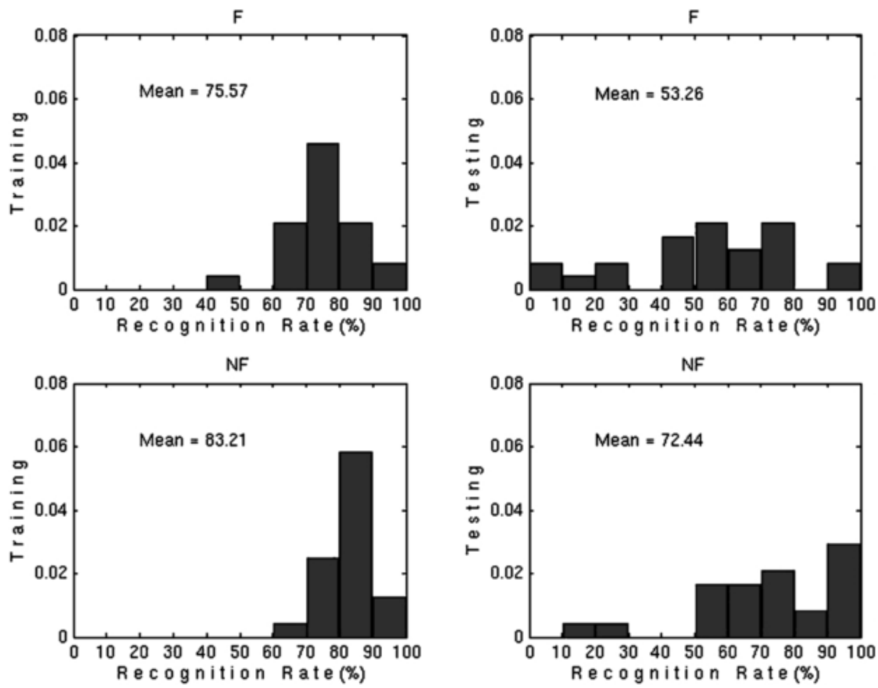

Figure 4.15: The classification rate for frustration in the top row. Bottom row is for no frustration. The top row is the classification rate for frustration when the samples are labeled frustration and the bottom row is the classification rate for no frustration when the samples are labeled no frustration. The results are from 24 participants. Plots on the left are classification rates for the training set and for testing set on the right.

Study	Class. classes	Chance level	Findings	PP diff.
Chanel 08 (boredom)	3	33%	72.5%	39.5
Chanel 08 (anxiety)	3	33%	68.4%	35.4
Lisetti 04 (KNN)	6	17%	72.3%	55.3
Lisetti 04 (DFA)	6	17%	75%	58
Lisetti 04 (MBP)	6	16%	84.1%	67.1
Kim 04	3	33%	78.4%	45.4
Kim 04	4	25 %	61.8%	36.8
Scheirer 02	2	50%	67.4%	17.4

Table 4.3: Comparison of the results of other physiological measures. PP is an abbrev. for percentage points

encouraging and provide good guidelines for a future implementation of the current project.
The other three studies are about classifying emotion based on physiological signals. All of the studies use stimuli which occupy two or more senses. For the Lisetti 04 study the stimuli was movie clips, the Kim 04 study used a mix of cognitive, visual and aural stimuli and the Scheirer 02 study used a visual puzzle. Another study is the Chanel 08 study described in the EEG chapter. Chanel used a computer game which involves cognitive and visual stimuli.
A comparison of the results is found in Table 4.3 below. Here the findings are much better compared to emotion classification using EEG signals. The worst result here (Scheirer 02) is still better than the best result using EEG signals.

The Lisetti 04 study presents the best results though one worry is that the classifiers might have been made to classify according to which clip the participants were watching instead of their emotion when watching the clip.
There are also differences in how the classification was done in the different studies that could impact the classification rates. Most notably if the test set was generated with samples from all participants the classifier is then trained for all participants and the classifier is less useful for predicting emotion for others. The Kim 04 study divided the participants into two groups: test and training and had some good results. The Scheirer 02 study trains a classifier for each participant which is not useful since each classifier only works for one participant. The Lisetti 04 study does not describe how the training set was generated while the Chanel 08 created a classifier for each participant which consisted of a training set which was composed of the other participants.
This is important when judging the results. The results from the Scheirer 02 study would not be usable for practical purposes because a classifier has to be trained for every new person. In comparison the Kim 04 classifier is plug and play – that is, it is usable on new people.

Study	Measures	Class. Method	PP diff.
Chanel 06 (2 class)	EEG	FDA	10
Chanel 06 (3 class)	EEG	FDA	12
Bos 06 (arousal)	EEG	FDA	15
Bos 06 (valence)	EEG	FDA	15
Chanel 08 (boredom)	GSR, BVP, HR, temp	RBF SVM	39.5
Chanel 08 (anxiety)	GSR, BVP, HR, temp	RBF SVM	35.4
Lisetti 04	GSR, HR, temp	KNN	55.3
Lisetti 04	GSR, HR, temp	DFA	58
Lisetti 04	GSR, HR, temp	MBP	67.1
Kim 04 (3 class)	GSR, HR, temp	SVM	45.4
Kim 04 (4 class)	GSR, HR, temp	SVM	36.8
Scheirer 02	GSR, BVP	HMM	17.4

Table 4.4: Comparison of the measures and classification methods. PP is an abbrev. for percentage points

Looking at the data analysis methods the different studies choose different methods. Table 4.4 summarizes the results and data analysis methods of the studies. The results show that the most successful studies uses both GSR, HR and temperature. This suggests that these three measures are very useful in recognizing emotion. Four of the studies use SVM and the results range from 35.4 to 45.4 which highly suggests that SVM is preferred. KNN, DFA and MBP show greater results but they are all from the Lisetti 04 study which says more about the study than the methods. The choice is between the MBP which is the best performing method and SVM which show great results across two studies.

The two EEG studies (Chanel 06 and Bos 06) furthermore trained classifiers for each participant while Chanel 08, Kim 04, Scheirer 02 and Lisetti 04 trained classifiers that were not user-specific which definitely is in favor other physiological measures and not EEG.

The mean of the difference in percentage points is 44.4 for physiological signals without EEG compared to the mean of 13 for EEG signals suggests that using a mixture of GSR, HR and skin temperature as physiological measures and classifying using SVM or MBP is the best solution for this project.

Chapter 5

Behavioral Measures

To create a classifier to recognize emotions it has to be trained first. Therefore there has to be a method for label the physiological samples with the emotion felt. In the section Physiological measures, a lot of the studies use the Game Experience Questionnaire or the Self-Assessment Manikin to establish the relation between a person's self-reported emotions and the physiological reaction of the same person. These two measures will be described below.

5.1 GEQ

GEQ or Game Experience Questionnaire is a questionnaire created by Poels, Kort and Ijsselsteijn [PKI07]. The GEQ was created because of the under-representation of game experience in academic circles. The GEQ provides a common framework for measuring game experiences.
The GEQ was created by using focus groups that were tasked to individually reflect on game experiences before a group discussion was held. Five game researchers were also invited to perform the same procedure as the focus groups. After the focus groups testing the results were gathered and Poels et al. combined them to form nine different game experience dimensions. The dimensions are shown in Figure 5.1.

Dimension	In-game experiences	Post-game experiences
ENJOYMENT	*fun, amusement, pleasure, relaxation*	*energised, satisfaction, relaxation*
FLOW	*concentration, absorption, detachment*	*jetlag, lost track of time, alienation*
IMAGINATIVE IMMERSION	*absorbed in the story, empathy, identification*	*returning to the real world*
SENSORY IMMERSION	*presence*	*returning to the real world*
SUSPENSE	*challenge, tension, pressure, hope, anxiety, thrill*	*release, relief, exhausted, euphoria*
COMPETENCE	*pride, euphoria, accomplishment*	*pride, euphoria, accomplishment, satisfaction*
NEGATIVE AFFECT	*frustration, disappointment, irritation, anger*	*regret, guilt, disappointment, anger, revenge*
CONTROL	*autonomy, power, freedom*	*power, status*
SOCIAL PRESENCE	*enjoyment with others, being connected with others, empathy, cooperation*	*accomplishment in a team, bonding*

Figure 5.1: Dimensions of game experience.

The study Drachen 10 [DN10] uses some of the dimensions to find correlates between game experience and GSR and HR and shows that elements of game experience does correlate with physiological measures.

5.2 SAM

The SAM or Self-Assessment Manikin [BL94] is a pictorial instrument to assess pleasure or valence, arousal and dominance in response to an object or event. The SAM was created to address issues with the Semantic Differential Scale which is also used to assess valence, arousal and dominance. The issues with the Semantic Differential Scale is that contains 18 ratings on a 9-point scale which is cumbersome to use if a participant would have to rate each stimulus on the scale. Another issue is that it is verbal e.g. the participants are therefore asked to rate whether a stimulus is unhappy to happy on a 9-point scale. This can cause problems in non-English studies due to translation or with children who might not understand the nuances of different emotions.
The SAM directly assess valence, arousal and dominance using a 9-point scale. The SAM can be seen in Figure . Since there are only 5 images on each dimension the participant can also choose to rate a stimulus as being in between the images. This leads to a 9-point scale.

Of the two the GEQ is the more game-specific. The problem with GEQ is that it has 9 dimensions compared to the SAM with 3 dimensions. This means that the GEQ has 3 times the dimensions of the SAM and if the game has to ask the player to fill out the GEQ 6 times then the player would have had to answer 45 questions compared to just 15 with the SAM. Apart from being annoying, it is also immersion-breaking to have to answer a large questionnaire 6 times in a row.
Furthermore using the arousal/valence model the SAM is much more suited since it measure arousal and valence directly. It even does it pictorially which more or less sidesteps the issues with different languages and cultures that are present in the GEQ.
The recommendation for the current project is to use SAM because it is less intrusive and therefore less immersion-breaking, directly measure arousal and valence and does not rely on definitions of words and concepts.

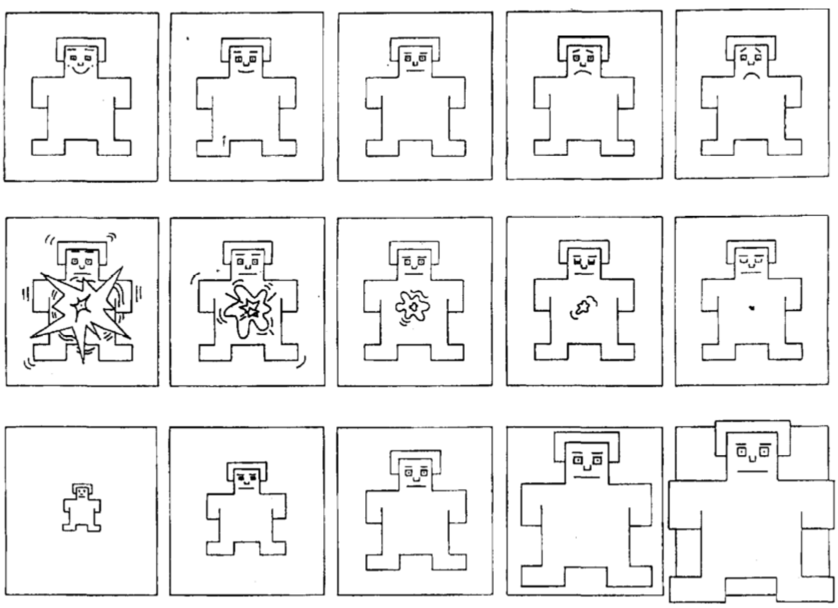

Figure 5.2: The SAM. The top row is valence, middle is arousal and bottom is dominance.

Chapter 6

Conclusion

The computer game industry has grown tremendously in the past years. The reason is that games provide people with entertainment and a sense of escapism. Most games are created to be fun although not all are. A system was proposed in the beginning of the report which could help improve the fun in games. This system is to measure the player's mental state and improve the game played to optimize the fun for the player. This system is composed of several parts. The first part is to establish how emotions are represented in the brain. Here Hamann [Ham12] proposes a network representation where each emotion activates several parts of the brain instead of each emotion corresponding to one part of the brain. This however is best measured with fMRI which is not a portable technology and therefore not suitable. Left is the dimensional theory where all emotions exist on a two dimensional map. This representation is chosen since the two dimensions, arousal and valence, can be measured with heart rate, facial muscles or skin conductance. These measures receive stronger signals than an EEG which could be used to measure emotions according to the other theory where an area of the brain equals an emotion.
Now that a view of emotions has been established as being dimensional, a clear view of fun in games must also be established. Fun in games can be compared to being in flow. The psychology of flow is a concept coined by Csikzentmihaly and defines a flow state as a state where a person is in complete focus of the task and feels a loss of sense of time, merging of awareness and action and a loss of self-consciousness. Csikzentmihaly has furthermore plotted flow in an abil-

ity/challenge map where great ability and low challenge will lead to boredom, great challenge and low ability will result in anxiety and equal challenge and ability will lead the player to be in a flow state.
Fun in a game can therefore be induced by leading the player into a flow state. Salen and Zimmerman [SZ04] discusses Dynamic Difficulty Adjustment (DDA) as a technique to dynamically regulate the level of challenge posed to a game player. A study by Hunicke [Hun05] and the use of DDA in commercial games show that DDA is a tool which can be used.
With the dimensional view of emotions and the concept of flow and DDA the emotions of the player has to be measured. Two methods are presented. One with EEG and measurement on the electrical signals of the brain and one with heart rate sensors, galvanic skin response sensors and other sensors which does not measure on the brain.
Comparing the results the EEG clearly lose to other physiological measures where EEG on average fared 13 percentage points better than chance level while other physiological measures were 44.4 percentage points better than chance level. Dissecting the performance of the other physiological measures show that galvanic skin response, heart rate and temperature were used by most studies and support vector machines were the most commonly used method for classifying the signals into emotions. This suggest that measuring galvanic skin response, heart rate and temperature and using support vector machines to classify emotions is the preferred method.
The measuring of emotions has to be classified using support vector machines. Since every person shows different patterns when reacting to events the game must include the training of the support vector machines. This can be done in the starting phase of the game and typically by asking the players how they feel. Two self-report schemes are useful. One is the game experience questionnaire which is a 9-dimensional scale where the dimensions relate to game experience and the other is the self-assessment manikin which is a pictorial 3-dimensional scale and relates to arousal, valence and dominance. The self-assessment manikin is preferred since it is faster to complete, is pictorial and completely relates to the chosen view of emotions as being on an arousal/valence map.

In the end the system is composed of physiological sensors that measure heart rate, galvanic skin response and temperature. The system will ship with a classifier that has already been trained. The training consists of players that will fill out self-assessment manikins while playing a game that will evoke different emotions. This data plus the signals from the sensors are used to train a support vector machine that can classify emotions based on arousal/valence. Based on studies the results should be around 44.4 percentage points better than chance level. When the game is played by a normal consumer the system will continuously measure the player emotions and if the player falls outside the flow zone the dynamic difficulty adjustment system can vary the challenge so that the player can reenter the flow zone and ultimately provide a fun and engaging game experience for the player.

Appendix A

Index of Terms

Arousal - Arousal is a measure of how alert a person is and in the valence/arousal dimensional emotion theory it ranges from calm to excited.
BVP - Blood Volume Pulse is a method to measure the blood flow using infrared light.
DDA - Dynamic Difficulty Adjustment. A technique to dynamically alter the difficulty in games.
Dominance - Dominance is an additional dimension that can be used with the valence/arousal dimensional emotion theory. It ranges from submissive to dominant.
EEG - Electroencephalography is the recording of electrical activity on the scalp.
EMG - Electromyography is the recording of electrical activity produced by muscles.
EOG - Electrooculography is a technique to measure the electricity used to move the retina. In reality it provides a measure on how the eye has moved.
Flow - A concept coined by Mihaly Csikszentmihalyi to represent a state of complete and energized focus in an activity with a high level of enjoyment and fulfillment.
GEQ - Game Experience Questionnaire. A questionnaire created by Poels, Kort and Ijsselsteijn [PKI07]. The questionnaire contains seven dimensions of gameplay: enjoyment, flow, imaginative immersion, sensory immersion, suspense, competence, negative affect, control and social presence.

GSR - Galvanic Skin Response. Other names include Skin Conductance Level and electrodermal activity. GSR is a technique to measure the conductivity of the skin through the sweat level of the skin.
HR - Heart Rate. Heart rate can be measured through a heart rate monitor but also with BVP or other methods.
IADS - International Affective Digital Sounds. A database compiled by Bradley and Lang containing sounds that are rated with arousal and valence values.
IAPS - International Affective Picture System. A database compiled by Bradley and Lang containing pictures that are rated with arousal and valence values.
SAM- Self-Assessment Manikin. A non-verbal pictorial assessment technique that directly measures the pleasure, arousal and dominance associated with a person's affective reaction to stimuli.
Valence - Valence is a measure of how pleasurable an event is to a person and in the valence/arousal dimensional emotion theory it ranges from highly positive to highly negative.

Bibliography

[BL94] MM Bradley and PJ Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental ...*, 25(I), 1994.

[Bos06] DO Bos. EEG-based emotion recognition: The influence of visual and auditory stimuli. *Emotion*, 2006.

[Che07] Jenova Chen. Flow in games (and everything else). *Communications of the ACM*, 50(4):31, April 2007.

[CKGP06] Guillaume Chanel, Julien Kronegg, Didier Grandjean, and Thierry Pun. Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals. *..., classification and security*, pages 530–537, 2006.

[CRBP08] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era - MindTrek '08*, page 13, 2008.

[DN10] Anders Drachen and LE Nacke. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. *... on Video Games*, 2010.

[ESA12] ESA. Essential facts about the computer and video game industry, 2012.

[Gau12] John Gaudiosi. New Reports Forecast Global Video Game Industry Will Reach $82 Billion By 2017, 2012.

[Ham12] Stephan Hamann. Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in cognitive sciences*, 16(9):458–66, September 2012.

[Hic11] Jon Hicks. Bethesda confirms no dynamic difficulty in Skyrim, 2011.

[Hun05] Robin Hunicke. The case for dynamic difficulty adjustment in games. *Proceedings of the 2005 ACM SIGCHI International ...*, pages 429–433, 2005.

[KBK04] K H Kim, S W Bang, and S R Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical & biological engineering & computing*, 42(3):419–27, May 2004.

[LN04] Christine Læ titia Lisetti and Fatma Nasoz. Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *EURASIP Journal on Advances in Signal Processing*, 2004(11):1672–1687, 2004.

[MIC06] RL Mandryk, KM Inkpen, and TW Calvert. Using psychophysiological techniques to measure user experience with entertainment technologies. *... & Information Technology*, 2005, 2006.

[MW13] Merriam-Webster. Merriam-Webster, 2013.

[PKI07] Karolien Poels, Yvonne De Kort, and W Ijsselsteijn. It is always a lot of fun!: exploring dimensions of digital game experience using focus group methodology. *... of the 2007 conference on Future Play*, pages 83–89, 2007.

[R* 12] R* K. Rockstar Game Tips: Difficulty Settings - Getting Started in Max Payne 3, 2012.

[RMP13] Boris Reuderink, C Mühl, and M Poel. Valence, arousal and dominance in the EEG during game play. *International Journal of Autonomous ...*, 2013.

[Sal09] Adam Saltsman. Game Changers: Dynamic Difficulty, 2009.

[SFKP02] Jocelyn Scheirer, Raul Fernandez, Jonathan Klein, and Rosalind W Picard. Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14(2):93–118, February 2002.

[Sie06] Joe Siegler. Prey Weekly Development Update #7, 2006.

[SZ04] Katie Salen and Eric Zimmerman. *Rules of Play: game design fundamentals*. MIT Press, 2004.

[Tol08] Josh Tolentino. Good Idea, Bad Idea: Dynamic Difficulty Adjustment, 2008.

[Uni11] University of Florida. THE CENTER FOR THE STUDY OF EMOTION AND ATTENTION, 2011.