# Predicting Players' Emotions from Game Telemetry

Ingibjörg Ósk Jónsdóttir

Thesis of 60 ECTS credits
**Master of Science (M.Sc.) in Software Engineering**

August 2017

# Predicting Players' Emotions from Game Telemetry

by

Ingibjörg Ósk Jónsdóttir

Thesis of 60 ECTS credits submitted to the School of Computer Science
at Reykjavík University in partial fulfillment
of the requirements for the degree of
**Master of Science (M.Sc.) in Software Engineering**

August 2017

Supervisor:

David James Thue, Supervisor
Assistant Professor, Reykjavík University, Iceland

Examiners:

Stephan Schiffel, Examiner
Assistant Professor, Reykjavík University, Iceland

Deepa Iyengar, Examiner
S.M. Brain and Cognitive Sciences, MIT, U.S.A

# Predicting Players' Emotions from Game Telemetry

Ingibjörg Ósk Jónsdóttir

August 2017

**Abstract**

In video games, interactive storytelling systems often use a player model to tailor the storyline to a player's preferences, personality, or skills. The player's in-game actions are often used as an input for the model, but their emotions can also offer useful knowledge. People often find it hard to describe their emotions and therefore, we aim to measure them through their physiological response. Since players are usually not equipped with such devices in their natural gaming environment, we seek to render the devices unnecessary by developing a method to predict a player's emotions from their in-game actions. Our method involves a user study where the player's actions are tracked and their physiological response is recorded. We then compute three emotion features (arousal, valence, and dominance) and train several machine learning algorithms to predict those features from the player's in-game actions. Our results show that our method can predict a player's emotion features from their in-game actions more accurately than the results of a uniform random predictor.

# Spáð fyrir um tilfinningar spilara út frá hegðun þeirra í tölvuleik

Ingibjörg Ósk Jónsdóttir

ágúst 2017

**Útdráttur**

Gagnvirk sögukerfi í tölvuleikjum notast oft við líkan af spilurum til þess að aðlaga söguþráðinn að áhugasviði þeirra, persónuleika eða færni. Hegðun spilara í tölvuleik er oft notuð sem vísir í gerð spilaralíkansins en tilfinningar þeirra geta einnig veitt gagnlegar upplýsingar. Fólki finnst of erfitt að lýsa tilfinningum sínum og þess vegna stefnum við að því að mæla þær með lífeðlisfræðilegri svörun spilaranna. Þar sem spilarar eru yfirleitt ekki búnir slíkum tækjum í sínu náttúrulega leikjaumhverfi miðum við að því að gera þau óþörf með því að þróa aðferð sem spáir fyrir um tilfinningar spilara út frá hegðun þeirra í tölvuleik. Aðferðin felur í sér notendarannsókn þar sem við rekjum aðgerðir spilaranna og mælum lífeðlisfræðilega svörun þeirra við spilun tölvuleiks. Við reiknum síðan gildi þriggja tilfinningalegra eiginleika (örvun, löð og styrkleika) og þjálfum nokkur vélnámsreiknirit til þess að spá fyrir um gildin út frá hegðun spilaranna. Niðurstöður okkar sýna að aðferðin okkar getur spáð fyrir um tilfinningalegra eiginleika spilara út frá hegðun þeirra í tölvuleik með meiri nákvæmni heldur en niðurstöður samræmdrar handahópsspár.

# Predicting Players’ Emotions from Game Telemetry

Ingibjörg Ósk Jónsdóttir

Thesis of 60 ECTS credits submitted to the School of Computer Science
at Reykjavík University in partial fulfillment of
the requirements for the degree of
**Master of Science (M.Sc.) in Software Engineering**

August 2017

Student:

Ingibjörg Ósk Jónsdóttir

Supervisor:

David James Thue

Examiners:

Stephan Schiffel

Deepa Iyengar

........................................................................
date

........................................................................................................................................................................
Ingibjörg Ósk Jónsdóttir
Master of Science

# Acknowledgements

I worked closely with Dr. Paolo Gargiulo, Associate Professor, at The Icelandic Center for Neurophysiology at Reykjavik University and I want to thank him for his help in shaping the EEG data acquisition protocol for this study and for allowing me full access to his laboratory. I also thank Dr. Kyle Edmunds for brainstorming with me about possible EEG setups in the study. I give my thanks to the Technical Operator, Patricia Edmunds, for training me in conducting the EEG study by myself.

The study received a very welcome reception and I got more participants for the study than I needed. I'm very thankful to all of the participants that took part in the study.

I want to thank my supervisor, David Thue, for his time, patience and guidance. I am very grateful for my mom, Kristín Ósk Hlynsdóttir, and I thank her for proofreading my work and inspiring me to pursue a master's degree. Last, but not least, I thank my fiancé Árni Freyr Haraldsson for standing by me like a rock and my family and friends for all their love and support.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| ANOVA | Analysis of Variance |
| BVP | Blood Volume Pressure |
| EEG | Electroencephalography |
| EMG | Electromyography |
| FFT | Fast Fourier Transform |
| GSR | Galvanic Skin Response |
| HCI | Human Computer Interaction |
| HRV | Heart Rate Variability |
| ICA | Independent Component Analysis |
| MSc | Masters of Science |

# Chapter 1

# Introduction

> You can't stay in your corner of the forest waiting for others to come to you. You have to go to them sometimes.
>
> *A.A. Milne, Winnie-the-Pooh*

Storytelling is an art form that has engaged humanity for tens of thousands of years. We have become addicted to storytelling as a species, as even in our sleep, our mind continues telling itself stories. Humans are perhaps the only animal that creates and tells stories, and thus it may be that storytelling is what makes us human [1]. Storytelling has taken on many different forms as humanity has developed. Some archeologists believe that even before we had developed writing, rock art was used as a form of storytelling [2]. Before modern technology, people would gather around and listen as a storyteller would use words and gestures to take their mind on a journey exploring a story world. As the story progressed, the audience would voice their opinions or questions which allowed the storyteller to learn about their interests and emotions, and adapt the story accordingly (e.g., by diving deeper into certain aspects of the story). Interactivity enabled the people to get even more invested in the stories as they took part in shaping them. For centuries, books have been a popular storytelling form, but most have omitted the interactive aspect of stories. In 1969, Packard wrote a series of interactive novels called *Adventures of You* which were later rebranded by Bantam Books as *Choose Your Own Adventure* [3]. The novels were formulated such that every few pages, the readers would have to choose between several options, leading them to different storylines. With modern computers, interactive storytelling has made its way into video games, which allows them to offer a personalised experience in which the player's actions influence how the story pans out.

In recent years, the study and development of interactive narrative experiences in video games has received increased attention. Intelligent storytelling systems enable video games to expand the typical one-way storyline and to create narratives that appeal to a greater number of players (e.g., through automatic personalisation [4]). This can be achieved by collecting data about the player through player models which can then be used to adjust the storyline according to the player's preferences, personality, or skills. Hence, players who prefer fighting may find themselves more often than not in a war, while players who prefer tactics can engage in meaningful conversations where hard decisions have to be made. Consequently, players are offered a potentially more enjoyable experience suited to their preferences. Various player modelling techniques have been used in the past to identify player preferences, with different player inputs being used to build a player model [5], [6]. The player's *narrative decisions* (i.e., a decision within a dialogue) in a video game offer a strong indication of their preferences. Other indicators such as *game telemetry* (i.e., a time-stamped history of the player's in-game actions) are also valuable. For example, if the player hesitates before taking an action, the action can have a smaller influence on the player model than if they had not hesitated [6].

By considering only the player's narrative decisions and telemetry for player modelling purposes, we risk missing valuable information. Interactive narrative video games are usually designed to offer an emotional experience and aim to allow the player to fully immerse themselves in a captivating storyline [7]. The player's narrative decisions and other in-game actions are driven by their thinking, and how they feel has a direct impact on what they think. The player's *feelings* during a narrative decision can, therefore, offer a valuable contribution to the player model. People often find it hard to control how they feel, since our *emotions* occur in our body without us realising it (e.g., we may feel an aching stomach without understanding it is caused by anxiety). Feelings and emotions have often been considered to be the same thing, but they are different [8]. Emotions are our physiological response and feelings are purely our interpretation of our emotions. Therefore, to evaluate a player's emotions, we cannot rely solely on self-assessed emotion questionnaires, since what they feel may not reflect their emotional state accurately. To assess the player's emotions, we must measure their physiological response. Since the player's actions are influenced by their feelings and their emotions power their feelings, we could potentially model the player more accurately by incorporating the player's emotions into the player model in addition to their narrative decisions and game telemetry.

Various techniques for measuring physiology can be used to estimate a player's emotions. A physiological response such as increased heart rate tells us the player is experiencing high arousal which could indicate excitement [9]. Techniques to evaluate a wider range of emotions than arousal include facial electromyography (EMG) and electroencephalography (EEG). Facial EMG records the electrical activity of facial muscles and can be used to extract the player's emotional features [10]. The technique measures facial EMG via electrodes that are positioned on the player's face. EEG is a data collection method which records the brain's electrical activity. The method involves an

EEG cap that must be placed on the player's head. The cap measures electrical activity through its electrodes which are filled with conductive gel. By applying feature extraction techniques to the recorded EEG signal, we can estimate emotion features including arousal, valence, and dominance. Arousal is the physiological state of being awake or excited. It can indicate emotional calmness or excitement [11]. Valence classifies emotions as positive or negative (e.g., sadness is a negative emotion). Dominance reflects the power of the emotion (i.e., how strongly we experience the emotion). These values can then be used to classify emotions such as excitement, anger, sadness, and calm (e.g., high arousal, negative valence, and high dominance classifies as anger) [12].

To record a physiological response (e.g., with EEG), the recording devices must be placed on the player's body. Generally, players do not wear such devices while playing commercial video games. Therefore, recording physiological response during gameplay can be intrusive and possibly reduce enjoyment [9]. We therefore cannot rely on being able to collect physiological data during actual gameplay in the player's natural gaming environment. To overcome the need for recording devices, could we use the player's game telemetry to predict their emotional state?

In this dissertation, we aim to answer the question: Can players' game telemetry be used to predict their emotions? We present a detailed methodology for doing so in this work, offering a strong base for future efforts to improve player experiences by predicting player emotions.

This dissertation is structured in the following way: We start by providing background knowledge needed to better understand our work. We discuss the challenges involved in this study and give criteria for its success. Then, we explore related work and present our proposed methodology for predicting player emotions. We evaluate our approach, and finally, we discuss our contribution and future work.

## 1.1 Background

The work we present in this dissertation involves several academic disciplines. It consists of neuroscience, psychophysiology, computer science, intelligent narrative technologies, and user experience. In particular, we draw upon the intersections of these disciplines as they occur in interactive storytelling, player modelling, and electroencephalography (EEG). In this section, we present the theoretical foundation upon which our work is built.

### 1.1.1 Interactive Storytelling in Video Games

Storytelling hasn't always been an important part of video games. Early video games had a simple user interface consisting of pixels which you could move around on the screen. The games had clear goals that could lead the player to victory, but a storyline was absent. *Pong*, a simulation of the game of table tennis, is one example [13]. The goal was clear, but no story was involved. Storytelling was later brought into video games with games like *Colossal Cave Adventure* [14]. The game was a simple story where the player used text commands to explore a cave which was thought to be filled with treasures. In the early 90's, the concept of interactive storytelling in video games started to form with Chris Crawford's *Dragon Speech* at the Game Developers Conference in 1992 [15]. Crawford claimed that the industry was recycling old ideas and informed them of his plan to focus on implementing interactive storytelling systems. This speech has been considered a turning point in the game industry as it started to explore the field of interactive storytelling.

*Façade* was considered a significant advance in interactive storytelling systems when it was released in 2006 [16]. The game consisted of an interactive drama in which the player interacted with a couple that was going through a rough patch in their relationship. The player had the ability to interact with the non-player characters (NPCs) via written language. Depending on the player, they could either mend the relationship or help destroy it. Façade's architecture consisted of a character authoring language and a drama manager. A drama manager is an AI agent that is used in many interactive storytelling systems. The drama manager knows everything about the game world and determines what should happen next in the storyline depending on different factors such as the player's knowledge or a player model.

Interactive storytelling systems can differ in the agency that is given to their players (i.e., how much control they have over the story and how much is simply an illusion of control). In systems with high player agency, the narratives are often branching stories where the storyline changes depending on the player's actions in the game world. One of the greatest challenges facing interactive storytelling systems is the burden of authoring. With branching storylines, every branch must be authored, which is expensive for game development. Studies in the field have presented various authoring tools that aim to solve the issue. The *Thespian* framework seeks to use automation to ease the programming burden of authoring by making it a creative exercise for non-technical authors [17]. The characters created with Thespian show great promise as they can generalise a script and respond to events that occur in a different order from the original script. The framework enables the reuse of characters and simple character reconstruction without altering story details. Fendt et al. suggested that the storyline might be kept linear while offering a similar amount of player agency as branching stories [18]. Successful commercial video games such as *L.A. Noire* [19] and *The Walking Dead* [20] provide a relatively linear storyline even though they achieve the illusion of high player agency. Those games may trick the player the first time they play the game, but the next time, the game could leave the players disappointed when they discover how little effect they have on the storyline.

### 1.1.2 Player Modelling

The game industry has changed a lot over the past 25 years. The spectrum of players has expanded from being an isolated group of young white males to a diverse group of people [21]. The task of modelling players has, therefore, become an even more challenging task. Player modelling techniques can be split into three categories: 1) subjective, 2) objective, and 3) gameplay-based [9].

Subjective methods involve asking the player about their game experience. One technique is to ask the players questions throughout the experience [4]. This is very likely to affect the player's immersion negatively by pulling them out of the experience repeatedly. Another technique is to ask the players questions at the end of the experience [5]. This method relies on memory rather than catching the player's thought at the desired moment in time. Both techniques constrain the experimenter to a formulated questionnaire in which wording may bias the answers or limit them to a set of predefined answers.

Objective methods involve measuring physiological responses from the player during gameplay, especially during events that are expected to have an emotional effect on the player. The physiological response can be obtained in various ways. Facial electromyography (EMG) records the electrical activity of facial muscles and can be used to extract the player's emotional features [10]. Galvanic skin response (GSR) measures the conductivity of the skin and has been shown to correlate with the subject's level of arousal [22]. Blood volume pressure (BVP) sensor is an optical sensor that measures the changes in blood volume [23]. Heart rate variability can be calculated from the BVP measurement which has been associated with valence [24], [25]. Increased heart rate indicates high arousal and can be measured with a heart rate monitor such as a heart rate wristband [9]. EEG records the brain's electrical activity and has been used to assess the emotional state of the player [26], [27]. These methods have several limitations. Methods recording electrical activity are sensitive to noise (e.g., in the case of EEG, the player's muscle movements and electrical devices used in the experiment). Moreover, the recording devices must be placed on the player's body. For this reason, the objective method can be intrusive and affect the player's experience negatively.

Gameplay-based methods involve obtaining data from the player's actions in the video game. This approach allows for real-time player modelling since data can be gathered during gameplay. The player's decisions in the game can then been used to predict the player's preferences [5]. Thue et al. proposed *PaSSAGE*, an interactive storytelling system that learns the player's style of play based on their in-game actions. The player's in-game actions play a major role in predicting the player's preferences. The actions can include the decision the player makes in a dialog with an NPC, the objects they choose to bear (e.g., a favourite sword), and their actions such as stealing. Thue et al. used the player's actions as a contributor to the weight of different player types. They categorised players as follows: Fighters, Power Gamers, Tacticians, Storytellers, and Method Actors. El-Nasr et al. took another approach to player modelling and based their model on the player's personality traits: reluctant hero, violent, self-interested, coward, truth-seeker [6]. Like Thue et al., they observed the player's in-game actions, but also added a temporal measurement into the player model which indicated if and how much the player hesitated in taking a particular action. This measurement reflected was used to estimate how confident the player felt about the chosen action [6].

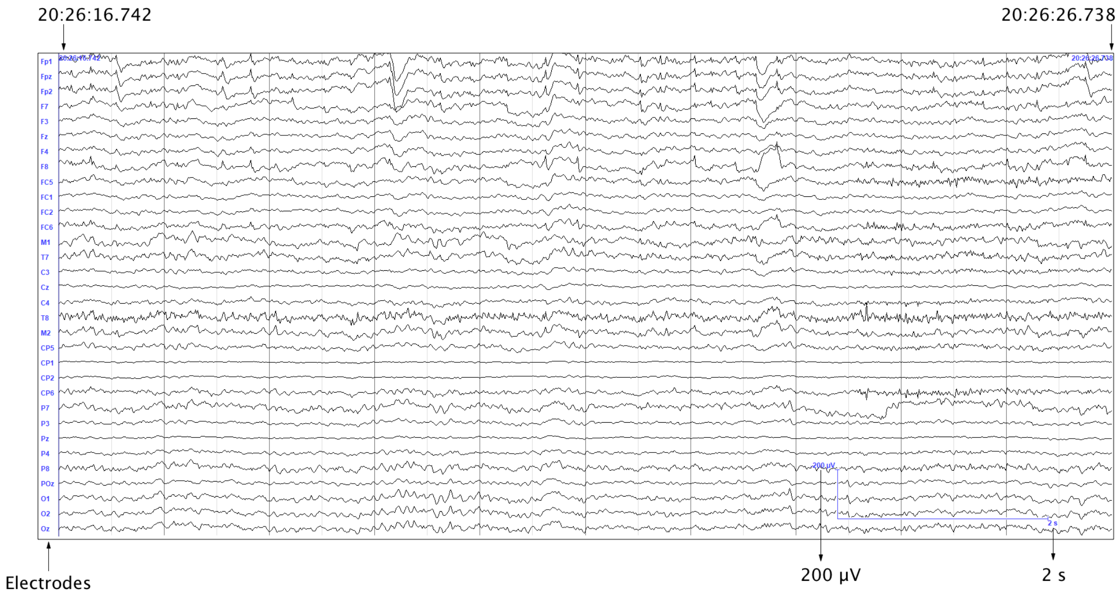



Figure 1.1: An example of an EEG recording with 32 electrodes.

### 1.1.3 Electroencephalography (EEG)

Psychophysiological methods have been applied successfully in studies of human computer interaction (HCI) [9]. The methods utilise physiological data to infer information about the subject's state of mind. One of these methods is EEG, which records electrical brain activity [28]. Since 1924, EEG has been used to record the electrical activity of the brain [29]. EEG has been studied in many different fields, which has resulted in its ability to differentiate between a wide range of psychological and physiological phenomena. EEG has been studied to a great extent for medical purposes, but in recent years it has attracted the interest of other fields, (e.g., the application of EEG in HCI [30]). In comparison with other recording methods, EEG is relatively non-intrusive. Recording of parameters such as blood volume pressure or galvanic skin resistance require the placements of sensors on hands, which restrict the natural hand movements of the player. While non-intrusive wristbands are available for optical detection of heart rate, they lack accuracy [31]. Facial EMG can provide a good measurement of a player's emotions, but the most accurate recording method requires placing sensors with wires on the face. We conducted this research at Reykjavik University which has a well equipped EEG laboratory. For these reasons, we used EEG to measure player's emotions. The plot in Figure 1.1 represents an EEG recording and shows the change in voltage during some amount of time.

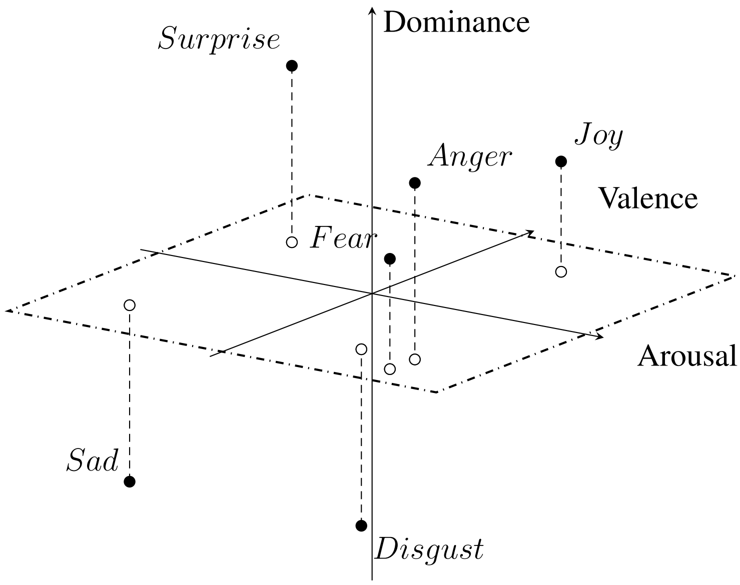



Figure 1.2: Positions of emotional states within an Arousal-Valence-Dominance plot. The black dot shows the position of the emotion and the white dot shows how it is positioned within the Arousal-Valence plane.

Emotion features have been extracted successfully from EEG recordings [12], [27], [32], [33]. The features extracted are values of arousal, valence, and dominance which can be mapped to certain emotions (see Figure 1.2). Moreover, EEG has been used to detect player emotions during gameplay [11], [26], [34]. Some studies have measured the accuracy of their methods by comparing the results to a self-assessed emotion questionnaire presented to the subject [35]. That is not entirely suitable, since the wording of the questionnaire might influence the answers. Also, many people aren't good at assessing their emotions. Presenting questions to the player during gameplay is likely to break the player's immersion. However, if the questionnaire is presented at the end of the gameplay, the player must rely on memory. Ramirez et al. presented their subjects with acoustic emotional stimuli sounds to evoke emotions and recorded an EEG signal. They did not verify their results with questionnaires but depended on emotional ratings (pleasure, arousal, dominance) provided with each sound from the International Affective Digitized Sounds (IADS) library [27]. Ramirez et al. proposed that their methods could be improved by incorporating self-assessed emotions into their classifiers.

| **Wave** | **Frequency Interval (Hz)** | **Origin** | **Properties** |
|---|---|---|---|
| $\delta$ | 0.5-3.5 | Thalamus | Deep sleep |
| $\theta$ | 3.5-7.5 | Hippocampus | Drowsiness and meditation |
| $\alpha$ | 7.5-12.5 | Thalamus | Working memory, alertness, focus |
| $\mu$ | 8-12 | Parietal lobe | Maintains physical stillness |
| $\beta$ | 12.5-30 | Parietal lobe | Busy thinking, anxiousness, concentration |
| $\gamma$ | 30-100+ | Hippocampus | Memory process, problem-solving, fear, and consciousness |

Table 1.1: Brain Wave Properties.

For our work, we aimed to extract features depicting the subject's emotional state during certain events in a video game. Our study focuses therefore on frequency analysis of an EEG signal and event-related potentials (ERP). Fast Fourier Transform (FFT) is a common frequency analysis algorithm in EEG analysis. An EEG signal over a certain amount of time is fed to the algorithm, which divides the signal into time blocks and outputs a frequency power spectrum [36]. The power spectrum is measured in micro volts squared ($\mu V^2$) for the brain waves delta, theta, alpha, mu, beta, and gamma (see Table 1.1). The alpha and beta waves are the most significant waves in regards to emotion feature extraction [27]. The alpha wave originates in the thalamus and captures information on alertness, the ability to focus, and working memory. The beta wave originates in the parietal lobe and has been associated with busy thinking, anxiousness, and concentration. Moreover, the activity in the beta wave relates to the salience of an emotionally evoking stimulus [28]. Studies have used results from FFT to successfully classify emotions from EEG signals [12], [27], [33], but none have tried to predict emotions from game telemetry.

# Summary

In this chapter, we described the art of storytelling and how it made its way into video games. We described the architecture of interactive storytelling systems and addressed the importance of using physiological response to model player emotions. We described three different player modelling techniques; subjective, objective, and gameplay-based. We introduced electroencephalography (EEG) as an option for objective player modelling which involves a physiological response from the players. As the technique involves placing a recording device on the player's body, we proposed to make the device redundant by predicting the player's emotions from game telemetry via physiological responses.

# Chapter 2

# Problem Formulation

We seek to develop a methodology that allows one to predict players' emotions based on their game telemetry. Interactive narrative video games often aim to offer an emotional experience in which decisions affecting the narrative must be made by the player. Therefore, we focus on these types of games. In this chapter, we list our requirements or such a methodology and present criteria for evaluating its success.

## 2.1 Player Modelling

Various player modelling techniques have been introduced in recent years. Both the player's narrative decisions and game telemetry are good indicators of the player's preferred experience, but what drives these actions is the player's emotions. By incorporating the extra information provided by emotional state into player modelling techniques, we can potentially provide a more compelling experience for the player.

A user experiment must be conducted to collect player data. Players will play an interactive narrative video game offering an emotional experience. To reduce bias, the players in the study must not have played the video game before. The game will preferably be easy to play such that it does not restrict us to participants that are experienced gamers. A demographics questionnaire asking the players about background information such as age and gaming experience can give us useful information during data evaluation. The player's narrative decisions and game telemetry must be recorded (e.g., with a mouse and key tracking software) so they can be associated with the player's physiological response that reflects their emotional state. The game telemetry data must be processed after the study, and valuable features must be extracted (e.g., which narrative decision the player chose at any point in time). To automatically predict the player's emotions from their game telemetry, a machine learning algorithm must be trained with game telemetry as input and emotional state as output. The time stamps of narrative decisions in the game must be recorded to synchronise the physiological response data and game telemetry.

## 2.2 Player Physiology

The player's emotional state can be assessed by recording their physiological response with devices such as EEG. The signal from the EEG can be used to classify arousal, valence, and dominance which can then be mapped to the player's emotions [12]. A data acquisition protocol must be in place such that a high-quality EEG signal can be achieved. Some EEG recording systems (including the one available to us) can only record data up to 45 minutes. Therefore, the period of gameplay cannot last longer than that. EEG is sensitive to noise such as muscle movements, and therefore, a video game on a PC will give us better data rather than one on an entertainment console. A baseline recording of the EEG must take place in which the signal is recorded when the player is relaxed. By doing that we can better analyse the highs and lows of the signal during gameplay by comparing it to the baseline. After the recording, the EEG signal must be filtered from artefacts such as noise from electrical devices and eye blinks. Further analysis must include feature extraction to allows us to calculate values of arousal, valence, and dominance. The range of these values can differ between subjects, and therefore, we must normalise them. That will enable us to make a cross-player comparison.

## 2.3 Emotion Prediction Method

Since players are usually not equipped with an EEG recording device while playing video games, we aim to render the device unnecessary during gameplay by using game telemetry to predict the player's emotions. To achieve this, we must use a machine learning algorithm that takes game telemetry as input and outputs values which can be mapped to emotions (e.g., arousal, valence, and dominance). We need to acquire the player's game telemetry and emotion features during narrative decisions in a video game, as we expect the player to experience strong emotions at those events. The features extracted from the game telemetry must be chosen wisely to ensure they contribute valuable information to the machine learning algorithm. We can present the player with a self-assessed emotion questionnaire to verify the results from our algorithm as suggested by Ramirez et al. [27]. The questionnaire should be presented to the player at the end of the gameplay, to preserve their immersion during gameplay. For a machine learning task such as this, it's helpful to acquire as much data as possible. However, setting up and recording EEG data is very time-consuming. A balance between time and data must be found.

## 2.4 Criteria for Success

For us to argue that our method is successful, the results must show that a player's physiological response can be predicted from their game telemetry. We will declare our methods successful if we achieve higher prediction accuracy than a uniform random predictor would produce. We will estimate our predictor's accuracy by training it on a subset of the data and then testing it on the rest.

## Summary

In this chapter, we described the methodological and practical challenges we face in predicting emotions from game telemetry. We outlined the basic requirements for our user study and noted that a data acquisition protocol must be in place for recording a high-quality EEG signal. We addressed the type of video game needed and how we must process the recorded data with game telemetry mining techniques to extract valuable features for a machine learning algorithm. We also discussed how we could obtain features from the EEG data, giving us values that we can associate with the player's emotions according to a certain model. Finally, we described how we would evaluate our methods to determine their success.

# Chapter 3

# Related Work

As mentioned in Section 1.1 our study involves several academic disciplines. In this chapter, we discuss prior work that has also studied the intersection of these disciplines. We start by presenting studies on emotion recognition with EEG. Then we review more highly related studies that have explored emotion recognition with EEG in video games.

## 3.1 Emotion Recognition with EEG

One of our goals in this study is to measure a player's emotions via EEG. Studies have found correlation between EEG and emotions [32], [37], [38]. Schmidt et al. explored whether the emotion features of valence and intensity could be distinguished within EEG data [32]. They were able to associate patterns of frontal brain activity with both features. They associated higher activity in the left frontal lobe to positive emotions, especially for the alpha wave, and a higher activity in the right frontal lobe to negative emotions. By examining patterns in the overall frontal brain activity, they were also able to distinguish the intensity of the emotion.

Davidson et al. observed facial behaviour and measured EEG while their subjects experienced the emotions, happiness, and disgust. One of their goals was to determine whether the emotions could be associated with patterns of hemispheric activation in different regions of the brain. They collected EEG data from 37 right-handed subjects and used short film clips as emotion stimuli that had been used in related emotion studies. The subject's facial responses were also recorded on a video. After each film showing, the subject was asked to rate the emotions they had experienced. Davidson et al. observed the video recordings after the experiment and rated the facial expressions with the Ekman-Friesen Facial Action Coding System [39]. They compared EEG recorded during a happy facial response and a disgusted one and were able to associate happiness with less alpha power in the left frontal region when compared to disgust and disgust to less alpha power in the right frontal region when compared to happiness. [37]. Davidson et al.'s findings have been utilised by other EEG emotion research [27], [33].

Baumgartner et al. studied the different effects that visual and musical stimuli have on emotions [33]. In addition to recording EEG, they also collected other physiological signals such as heart rate, skin conductance and more. Their methods involved collecting data from twenty-four subjects while they listened to music and observed images that were designed to evoke either fear, happiness or sadness. At the end of each session, the subjects were asked to answer two questionnaires regarding their emotional state. After collecting the data, they performed an FFT analysis on the EEG signal to extract power density values for the alpha wave only, as it has been reported to have a stronger correlation to behaviour than the other waves [40]. They analysed the alpha power density values with analysis of variance (ANOVA) using four factors: emotion (fear, happiness, sadness), modality (combined, visual, musical), region (anterior, posterior) and hemisphere (left, right). The results showed higher accuracy in emotion recognition when the modality was combined (i.e., both visual and musical stimuli).

Other studies have used a machine learning algorithm to learn patterns related to emotions. Thammasan et al. conducted a study where they observed whether familiarity affected an EEG signal, especially in EEG-based emotion classification systems [41]. They collected data from a WaveGuard EEG cap while subjects listened to eight familiar songs and eight unfamiliar songs. To determine whether familiarity affected the EEG signal, they performed two different analyses: single-electrode-level power spectral density analysis and functional connectivity analysis. They found in both cases that familiarity does affect the EEG signal. Then, they observed the effects of familiarity in an EEG-based emotion recognition system. To extract features from the EEG signal and estimate the emotional state, they observed the fractal dimension value and power spectral density values. To classify emotions, they used Russell's arousal-valence emotion model [42] and classified the values of arousal and valence into two classes, high and low. Finally, they applied three classifiers on their data to classify values of arousal and valence. Their results demonstrated a better emotion classification on data collected when the subject listened to an unfamiliar song than to a familiar one. These results support only including players who have not played the video game used in our study.

Using machine learning algorithms as well, Ramirez et al. applied two classifiers on an EEG recording they obtained from six participants using consumer mass-market grade equipment [27]. The participants listened to twelve sounds provided by the *IADS* library that had already been classified with certain emotions. After filtering the EEG signal, Ramirez et al. performed an FFT analysis to extract values for calculating the arousal and valence. They measured arousal by computing the ratio between alpha and beta values in electrodes in the prefrontal cortex. To calculate valence, they compared the activation levels of the two cortical hemispheres since left frontal inactivation indicates a negative emotion, and right frontal inactivation a positive one as proposed by Davidson et al. [43]. They then trained and tested two classifiers which, given the EEG data at a particular point in time, were meant to predict arousal as one of two classes (high or low) and valence as one of two other classes (positive or negative). The classifiers were able to classify arousal with 83.35% accuracy and valence with 86.33% accuracy. Compared to a uniform random predictor which correctly classifies 50% instances on average, they were able to show that their methods can link EEG patterns to emotion features better than a random predictor. In our study, we use the same methods for calculating arousal and valence. Unlike Ramirez et al., we also want to predict the emotion feature dominance and our stimulus is also very different as we use events in a video game to evoke emotions.

To calculate the value of dominance, we used the same equation as Blaiech et al. [12], which was originally formulated by Liu et al. [44]. Blaiech et al. aimed to implement an emotion recognition system by inducing emotions with methods such as biographical reminders of the subject's memories or social interactions. They conducted a user study and obtained EEG data from six participants while inducing different emotional states with various methods. They based their stimuli in various ways. They induced sadness by asking the subject to recall an unhappy memory of theirs and they induced disgust by showing a video tailored to trigger disgust. At the end of the study, the participants answered a self-assessed emotion questionnaire to evaluate their emotional state. Blaiech et al. then filtered the EEG signals, conducted an FFT analysis, and computed the values of arousal, valence, and dominance. Finally, they used fuzzy logic techniques to classify each emotion by using values of arousal, valence, and dominance as input. They separated the values of each emotion feature into three categories of small, medium, and large, like we do in our study. Their results showed satisfactory classification of seven different emotions with the highest accuracy of 100% for neutral emotion state and the lowest accuracy for a surprised emotion at 53.57%.

The studies by Ramirez et al. and Blaiech et al. both collected EEG data using the Emotiv EPOC cap [45], developed for the consumer mass-market. The cap consists of fourteen electrodes, whereas the cap we used had thirty-two electrodes. The Emotiv EPOC cap has also been used in brain-computer interface studies [46], [47].

## 3.2 Emotion Prediction in Video Games

In our study, we aim to predict emotions during gameplay. Several studies have used EEG to compute a player's emotional state during gameplay [11], [48], [49]. Rodriguez et al. conducted an exploratory study on emotion recognition with EEG during gameplay [11]. Using consumer-grade equipment, they obtained EEG data from thirty participants and focused on extracting EEG data from different events within the game. After analysing the EEG data, they labeled the data with results from self-assessed emotion questionnaire which the players answered at the end of each game event. Rodriguez et al. classified the emotions into four different types - one for each combination of high or low levels of arousal and valence. For extracting the emotion features, they used both FFT analysis and calculated asymmetry index to build feature vectors which represented the EEG data for each game event. They then trained and tested several classifiers with data on emotions and the game's event types. They achieved a best accuracy of 41.45% in classifying four game events and 33.48% in classifying the combinations of levels of arousal and valence. Rodriguez et al. concluded that they could have possibly gotten better results by omitting the repeated questionnaires to preserve the player's immersion. In our study, we take this recommendation seriously and only present the participants with a self-assessed emotion questionnaire at the end of their gameplay.

Chen et al. were aware of the importance of preserving the player's immersion in their study of relationships between game events and a player's emotions [48]. They hypothesised that emotion changes during gameplay are consistently triggered by game events. The study recorded EEG signals for twenty players while they played three types of video games. In addition to EEG, Chen et al. also recorded the player's facial expression and their computer screen for syncing the game events together with the EEG data. For each game, they annotated different game events with two different emotions: excitement and frustration (e.g., when hitting a target while playing the video game *Battlefield 4*, the player would experience excitement, but frustration when they would die or fail a mission). Chen et al. found a strong correlation between game events and peaks within the EEG data. They concluded that games could be adapted to improve player satisfaction by using emotion data triggered by game events. For adapting video games in real-time, their proposed approach requires EEG data collection during gameplay, but we wish to avoid that because players do not typically have access to such devices. In Chen et al.'s study, they annotated the emotions themselves; but in our study, we aim to compute them with the equations mentioned in Section 3.2. However, we utilise the fact that game events trigger emotions in our study by analysing EEG data for each game event separately - not the whole duration of the gameplay.

Game adaption based on emotions was also explored by Chanel et al. [49]. They aimed to adapt the difficulty in video games by using values of boredom, engagement, and anxiety. Their study acquired data from twenty players while they played the video game *Tetris* several times at three levels of difficulty: easy (boredom), medium (engagement), and hard (anxiety). In addition to EEG, they used several devices to measure physiological response to classify emotions. They measured skin resistance, blood volume pressure, the temperature of the subjects' palms, and extension of the abdomen. Like Rodriguez et al., they asked the player to answer a self-assessed emotion questionnaire at the end of each session. After filtering the EEG signal, they conducted a feature extraction by using an FFT analysis. Then, they trained and tested three classifiers to predict one of the three emotional states: boredom (easy), medium (engagement), and hard (anxiety), from the physiological response. The best performance for using EEG as input was 56%, but it was 59% with the other signals. Combining the results of the signals improved the performance and accuracy was reported to be 63%. Chanel et al. claimed that EEG is better suitable for short-term emotion assessment. We utilise this fact in our study by only analysing five seconds of EEG data at a time, in accordance with each game event. Chanel et al.'s way of asking the player questions between sessions may have resulted in the player becoming more aware of their emotions during gameplay which possibly affected their experience. Similarly to Chanel et al.'s work, this study requires an EEG device during real gameplay, but we seek to use game telemetry to predict the emotions such that we don't need to rely on players owning EEG recording devices.

Derbali et al. used EEG to predict the player's motivational state during gameplay [50]. Their methods included a user study of thirty-three participants who played a game called *FoodForce*. The goal of the game is to teach the player about world hunger, and the player must feed as many people as possible with a low budget. Derbali et al. used two cameras to record the player's facial expressions and their in-game actions. They also measured Galvanic Skin Response and Blood Volume Pulse with a sensor which was attached to the player's finger, leaving one hand free for interaction with the computer. Before the play session, the researchers recorded a baseline from the physiological response in which the player relaxed with their eyes closed for 60 seconds. At the end of each session in the game, the player was asked to answer a questionnaire designed to identify four components of motivation: attention, relevance, confidence, and satisfaction. At the end of the study, the participants answered the questionnaire again such that a final motivation value could be obtained. Like in the other EEG studies we've discussed, Derbali et al. conducted an FFT analysis on the EEG data. Each 60-second section of the EEG data corresponding to a mission in the game was analysed. Then, they normalised the FFT values and used them to predict a player's motivation. They found that the theta wave in the frontal region and the high-beta wave in the left central region provided a predictor for the player's motivation. In our study, we also record a baseline and normalise the FFT values. However, we do not aim to predict the player's motivational state, but instead the emotion features arousal, dominance, and valence.

None of the studies address the fact that we can usually not acquire a player's physiological response in their natural gaming environment. To the best of our knowledge, the goal of our study to predict players' emotions from their game telemetry has not been explored before.

# Summary

In this chapter, we discussed the two studies on which we base our arousal, valence, and dominance equations (see Section 3.1). We also presented other work that has aimed to use EEG to recognise emotions during gameplay. Finally, we reviewed work more highly related to our study, which aimed to predict emotions during gameplay.

# Chapter 4

# Proposed Approach

Our proposed methodology for predicting players' emotions based on their in-game telemetry involves three steps: data acquisition, feature extraction, and training and testing a machine learning algorithm on our data.

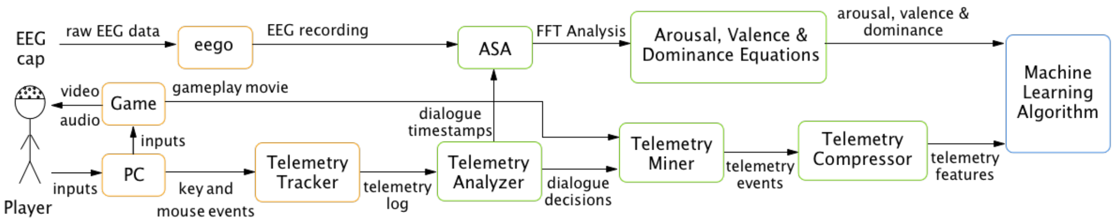



Figure 4.1: Schematic diagram of our emotion prediction methodology showing the three steps: data acquisition (orange), feature extraction (green) and machine learning (blue).

To acquire data, we conducted several play sessions in which human participants played through half an hour of an interactive storytelling video game called *The Wolf Among Us* by *Telltale Games* [51]. As shown in the schematic diagram given by Figure 4.1, we recorded the players' physiological response with an EEG cap and tracked their mouse and keyboard events using a telemetry tracking tool [52]. The latter allowed us to determine the players' game telemetry. The feature extraction step involved analysing the game telemetry and extracting features about the players' actions that would be valuable to a machine learning algorithm that predicts the players' emotions. This part also included processing the acquired EEG data and extracting the emotion features arousal, valence, and dominance. Finally, we trained several machine learning algorithms to predict the values of arousal, valence, and dominance from game telemetry.

In this chapter, we explain how we acquired the data required for training and testing the machine learning algorithm (Section 4.2). We provide a detailed description of the features we extracted from the game telemetry and the EEG data (Section 4.3), and then present our emotion prediction methods (Section 4.4). We begin with a brief introduction to the game from which we acquired our data.


Figure 4.2: Game Events in *The Wolf Among Us* [51]. Top: Dialogue event with a red timer bar below. Bottom left: Exploration event. Bottom right: Action event.

## 4.1 The Video Game

To obtain an emotional EEG recording, we needed to choose a video game offering an emotional experience. We decided to use the first episode of the interactive storytelling game *The Wolf Among Us* by *Telltale Games* [51]. The game has received "Overwhelmingly Positive" reviews on the entertainment platform *Steam* [53], and is designed to offer an evocative experience. It is structured like a television show in which the player plays the main character and can influence the storyline by making decisions on the character's behalf. By choosing this game, we made sure that each player would be presented with emotionally evoking content. It also allowed us to include inexperienced game players in the study since it is relatively easy to play. The only controls are the mouse and the traditional keyboard game scheme using the WASD (and Q) keys to move around. The player interacts with different NPCs throughout the game and can explore the game world (see Figure 4.2). We extracted features from both the recorded game telemetry and the EEG.

The video game consists of three kinds of interactive events, as shown in Figure 4.2. One is a dialogue event with an NPC. In this event, the player chooses between two or four different responses by clicking the desired response within a given time limit ranging from 5-10 seconds (Figure 4.2, Top). If the timer expires, the player's character gives a silent response, and the dialogue continues. Another kind of game event is the exploration event in which the player has unlimited time to explore some aspects of the game world (Figure 4.2, Bottom left). The player can walk around using the keys W, A, S, D, and use the mouse to take actions (e.g., open a door). The third kind of game event is when the player is instructed to move quickly in certain directions using the keys W, A, S, D or to use force by hitting the key Q repeatedly (Figure 4.2, Bottom right). In this study, we focused on the dialogue events since the other two included keyboard events which were more likely to contaminate the EEG signal with muscle artefacts. Each player in the experiment went through the same twenty-eight dialogue events.

## 4.2 Data Acquisition

We collected data via several play sessions in a laboratory used for HCI research at Reykjavik University. We conducted a pilot study with two participants to help shape the design and scheduling of the data collection. In addition to the two participants of the pilot study, twenty others partook in the play sessions. The participants all had sufficient computer skills and had a median age range of 25-29 years. We recruited participants via advertisements placed at the student union of those studying computer science at Reykjavik University, the student community at Reykjavik Academy of Digital Entertainment, and a local software company.

We divided the data acquisition into three parts; a demographics questionnaire, data collection from EEG and game telemetry during gameplay, and a self-assessed emotions questionnaire. We acquired the players' background information regarding their gender, age, etc. with a demographic questionnaire (see Questionnaire A.1). We recorded the EEG data with a CA-201 WaveGuard EEG cap. The cap consists of 32 electrodes and two reference electrodes. The electrode placement scheme is an extension of the 10/10 system proposed by the American Clinical Neurophysiology Society [54]. The electrode locations are Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, M1, T7, C3, Cz, C4, T8, M2, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, POz, O1, Oz, and O2 (see Figure 4.3) [55]. The EEG data was transferred from the cap through an amplifier to a tablet. This setup prevents any data loss since the amplifier and tablet are connected directly to one another, and do not rely on WiFi or other wireless connections.

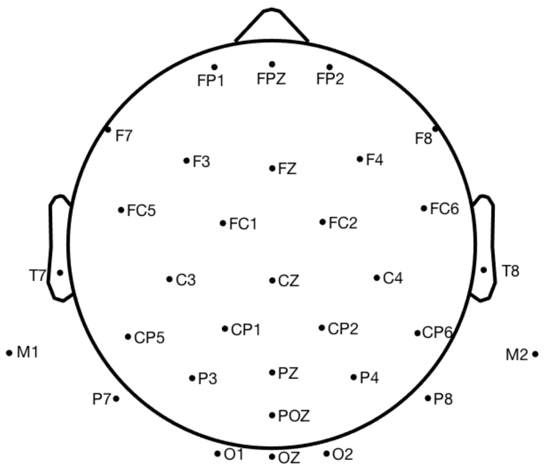



Figure 4.3: Cap Layout of 32 Channel WaveGuard EEG.

For machine learning tasks such as ours, it would be ideal to acquire as much data as possible. However, recording EEG data and preparing to do so is very time-consuming. Light et al. formulated a base protocol for setting up EEG studies in which they stated that standard EEG studies have 16-20 participants for each group of interest [56]. They suggested that each study should also review the related literature to decide on a number of participants. Since most of the EEG studies we discuss in Chapter 3 used twenty participants in their data acquisition [26], [48], [49], we decided to also obtain data from twenty players.

The device that we used for gathering the EEG data is only capable of recording for less than 45 minutes at a time. Therefore, we asked the players to play the first 30 minutes of the game, stopping them all at the same point in time. Up until that point, each player went through the same twenty-eight dialogue events from which we acquired the game telemetry. We left the players alone in the laboratory during gameplay to prevent any interference with their immersion (see Figure 4.4). We collected game telemetry with our custom built key and mouse tracking software called *Telemetry Tracker*. We also recorded the computer screen during gameplay with the open source video recording software *OBS Studio* [57] to allow for verification with the analysed mouse tracking data; we verified the time stamps of the dialogue events by comparing the times of both the mouse clicks in the telemetry and the relevant frames in the video. After playing, the participant self-assessed their emotions during the game by answering a Positive and Negative Affect Schedule Questionnaire (PANAS) [58] related to every scene in the game, where each scene consisted of several dialogue events (see Questionnaire A.2). The total duration of the data collection session was roughly one and a half hours, as we describe in Section 4.2.2.

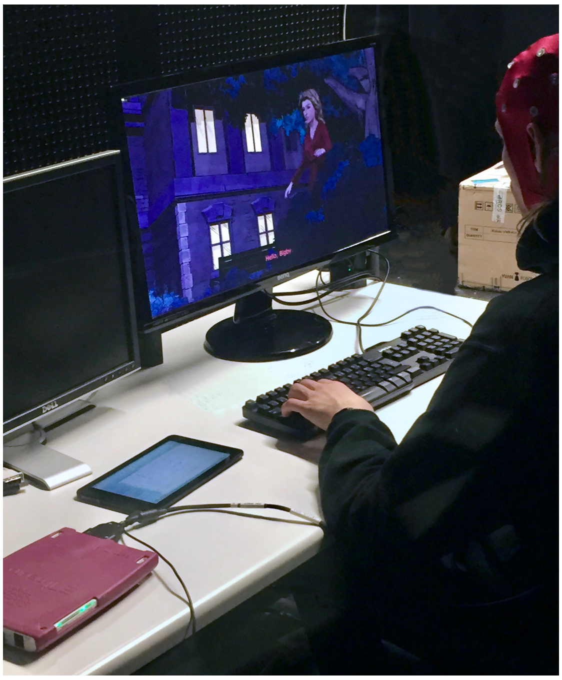


Figure 4.4: A participant playing the video game *The Wolf Among Us* while wearing an EEG cap. The pink device on the left is an amplifier that connects the EEG cap to the tablet.

### 4.2.1 Telemetry Tracker

For tracking the player's game telemetry, we adapted the *Mouse'n'Key* tracker tool [52], written in Python 3. The tool tracks and saves mouse and keyboard input on a Windows machine by using the *PyHook* package [59]. We developed the *Telemetry Tracker* [60] by altering the *Mouse'n'Key* tool to track the mouse input of interest and to log the desired values to a file. The values from each mouse event included a time stamp, cursor coordinates, and the type of mouse event (i.e., a mouse movement or a mouse click). We also recorded each keyboard event and logged it to a file.

### 4.2.2 Data Acquisition Protocol

The following list represents the protocol that we followed while acquiring data from players during gameplay.

1. **Introduction & Demographic Questionnaire (10 minutes)**
   The participant was escorted into the laboratory and asked to sit on a chair facing a computer screen. The researchers started by giving a detailed explanation of the EEG cap and the data collection procedure. After receiving the participant's consent for the session, the EEG cap was then placed on their head, making sure they were comfortable and that the electrodes were correctly laid down. While the EEG cap was being connected with the amplifier to a tablet and set up for recording in the software *eego*, the participant was asked to answer a demographics questionnaire regarding their age, etc. (see Questionnaire A.1). The EEG cap was set to record at a 512 Hz sampling rate.

2. **EEG gel insertion and game instructions (30 minutes)**
   The gel was inserted into the 32 electrodes on the EEG cap. Impedance was measured through *eego* and the researchers made sure it was below 100 k$\Omega$ to ensure a high-quality EEG signal. Meanwhile, the game controls were explained thoroughly to the participant.

3. **Baseline recording (1 minute)**
   A baseline recording was acquired to obtain the participant's brain activity in a relaxed state. EEG data was recorded while the participant relaxed with their eyes open for 30 seconds. Another recording was done with eyes closed for another 30 seconds [61].

4. **Analytics setup and take off (5 minutes)**
   The *Telemetry Tracker* and a screen recording were started. The participant was advised to keep their head as still as possible while playing to minimise any muscle artefacts in the EEG data. The EEG recording and the video game was started, and the researchers left the room, allowing the participant to immerse fully into the game.

5. **Gameplay and emotion questionnaire (45 minutes)**
   The participant played through approximately half of the first episode of *The Wolf Among Us*. Every participant played until they reached the same, predetermined point in the game, which took them 32 minutes on average. The researcher then entered the room and stopped the game and all recording devices. The participant was then asked to fill out a self-assessed emotion questionnaire for each scene in the game (see Questionnaire A.2).

6. **Head model taken with *Xensor* (5 minutes)**
   A head model was made with the software *Xensor* [62] to obtain accurate information about the electrode positions in 3D.

7. **Farewell (20 minutes)**
   The cap was removed and the participant was offered a towel and access to a bathroom to clean the gel from their hair. Finally, they were thanked for their participation, and the EEG cap was washed and dried.

### 4.2.3 Sample of Collected Data

The data files below show samples of mouse and keyboard data collected from the *Telemetry Tracker*. A sample of EEG data is shown in Figure 1.1.

Mouse Events

```
event_type;pos_x;pos_y;time
mouse_move;1055;929;1487877919618
mouse_move;1054;933;1487877919626
mouse_move;1054;935;1487877919634
mouse_move;1053;937;1487877919642
mouse_move;1053;937;1487877919651
[...]
```

Each mouse event in the sample data above includes the following information:

- *event_type*: The type of mouse event (i.e., mouse movement or mouse click).
- *pos_x*: The x (horizontal) coordinate of the cursor.
- *pos_y*: The y (vertical) coordinate of the cursor.
- *time*: The time stamp in milliseconds since Jan 1, 1970 00:00:00 UTC.

Keyboard Events

```
event_type;key_code;key_code_readable;scan_code;alt_pressed;time
key_up;65;A;30;False;1487878336874
key_down;65;A;30;False;1487878336938
key_up;65;A;30;False;1487878337018
key_down;65;A;30;False;1487878337098
key_up;65;A;30;False;1487878337242
[...]
```

Each keyboard event in the sample data includes the following information:

- *event_type*: The type of keyboard event (i.e., key down or up).
- *key_code*: The ASCII value of the key.
- *key_code_readable*: The character corresponding to the ASCII value of the key.
- *scan_code*: The keyboard scan code of the key.
- *alt_pressed*: A boolean value indicating whether the key *alt* was pressed down.
- *time*: The time stamp in milliseconds.

## 4.3 Feature Extraction

In the following subsections, we describe how we processed the data and which features we extracted from it to use in our emotion prediction method.

### 4.3.1 Game Telemetry

To obtain useful input for a machine learning algorithm that predicts the emotions, we processed the data from the *Telemetry Tracker* with three custom built telemetry analysis tools: the *Telemetry Analyzer*, the *Telemetry Miner*, and the *Telemetry Compressor* [60].

We developed the *Telemetry Analyzer* to analyse the mouse tracking data shown in Section 4.2.3. The *Telemetry Analyzer* provided us with the decisions the player made during each dialogue event in the game and the time at which each decision was made. In developing the *Telemetry Analyzer*, we needed to know the time between dialogue events because collecting the mouse clicks alone would give inaccurate data; the player might use the mouse in other game events or even double click on a dialogue event. Unfortunately, time intervals differed between dialogues depending on the chosen response, and we were therefore unable to know the exact time intervals between dialogue events for each possible case. Therefore, it was necessary to watch the screen recordings from each player to verify the output from the *Telemetry Analyzer*. The following sample shows data acquired after running the *Telemetry Analyzer* on the mouse tracking data shown in Section 4.2.3. As as illustrated example, we have made the first dialogue decision in the sample bold and we will continue to show how this sample changes in the next steps of the analysis by making the corresponding instance bold as well. The highlighted decision represents the player's response to an event in which they must choose to confront or forgive one of the story's characters (see Figure 4.2: Top).

Dialogue Decisions - *Telemetry Analyzer* Output

```
event_type;dialogue_decision;time
dialogue;I'm looking at a 3 foot toad.;1488718608099
dialogue;Get it fixed.;1488718647355
dialogue;Silence...;1488718670019
dialogue;Alright, why'd you hit her?;1488718784819
dialogue;[threaten him];1488718795699
dialogue;Say that word again.;1488718934531
[...]
```

Each dialogue event in the sample data above includes the following information:

- *event_type*: The type of event which was always *dialogue* in this case.
- *key_code*: The chosen dialogue response.
- *time*: The time stamp in milliseconds.

Once we had acquired the player's decisions and their time stamps, we developed the *Telemetry Miner* to extract every mouse tracking event for five seconds up until the player made a dialogue decision, for every such decision. We decided to use five seconds because although the player had a ten second window to chose a response in most of the dialogue events, in two cases the window was only 5 seconds. Also, the typical duration of emotions is approximately 0.5 to 4 seconds [63]. The *Telemetry Miner* provides an array of all mouse events up until the decision for each dialogue event, within the five seconds prior to the decision. Each mouse event includes eight values: time stamp, coordinates of the cursor, mouse event type (click or movement), and four boolean values (overA, overB, overC, and overD) which indicate over which dialogue response the mouse was positioned (see Figure 4.5).

Mouse Events per Dialogue - *Telemetry Miner* Output

```
[[0, 960, 540, 0, 0, 0, 0, 1488718605443],
 [0, 960, 540, 0, 0, 0, 0, 1488718605451],
 [0, 962, 540, 0, 0, 0, 0, 1488718605459],
 [0, 963, 541, 0, 0, 0, 0, 1488718605467],
 [0, 964, 542, 0, 0, 0, 0, 1488718605475],
 [...],
 [1, 752, 839, 1, 0, 0, 0, 1488718608099]]
[...]
```

The sample with mouse events per dialogue shows the output from the *Telemetry Miner*, which consists of an array of mouse events for each dialogue event. Each mouse event includes the following information:

- *click*: A boolean value stating whether the event type was a mouse click.
- *x*: The x coordinate of the cursor.
- *y*: The y coordinate of the cursor.
- *overA*: A boolean value stating whether cursor was over area *A*.
- *overB*: A boolean value stating whether cursor was over area *B*.
- *overC*: A boolean value stating whether cursor was over area *C*.
- *overD*: A boolean value stating whether cursor was over area *D*.
- *time*: The time stamp in milliseconds.


Figure 4.5: Dialogue options in *The Wolf Among Us* with annotations (A, B, C, D, and E).

We used the *Telemetry Compressor* to compress the output from the *Telemetry Miner* into an array of features. The array consisted of twenty-four game telemetry features for each dialogue event. For each response box and the area outside of the boxes, we stated whether the area was visited or not, the total time spent in the area, the total mouse velocity, and the visit frequency (i.e., how many times the player visited the area in total). The array also included information on which box, if any, was selected (i.e., top left, top right, bottom left, bottom right, or none, noted as A, B, C, D, or E in Figure 4.5). We also noted the total velocity of the mouse cursor before the player made a dialogue choice and the total visit frequency to each area. Finally, we included the total time before the player started enacting a dialogue decision (i.e., the time before the player started moving the mouse to select a response).

Final Game Telemetry - *Telemetry Compressor Output*

```
[1, 464, 5, 1, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 1, 8, 51, 1, 65, 56, 2, 8]
[1, 1985, 24, 2, 1, 360, 10, 1, 1, 680, 5, 1,
 0, 0, 0, 0, 1, 8, 6, 3, 66, 45, 7, 8]
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 1, 0, 0, 1, 83, 0, 1, 0]
[0, 0, 0, 0, 1, 416, 8, 1, 0, 0, 0, 0,
 0, 0, 0, 0, 1, 8, 84, 1, 66, 92, 2, 8]
[1, 472, 7, 1, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 1, 8, 51, 1, 65, 59, 2, 16]
[1, 345, 5, 1, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 1, 8, 5, 1, 65, 9, 2, 24]
[...]
```

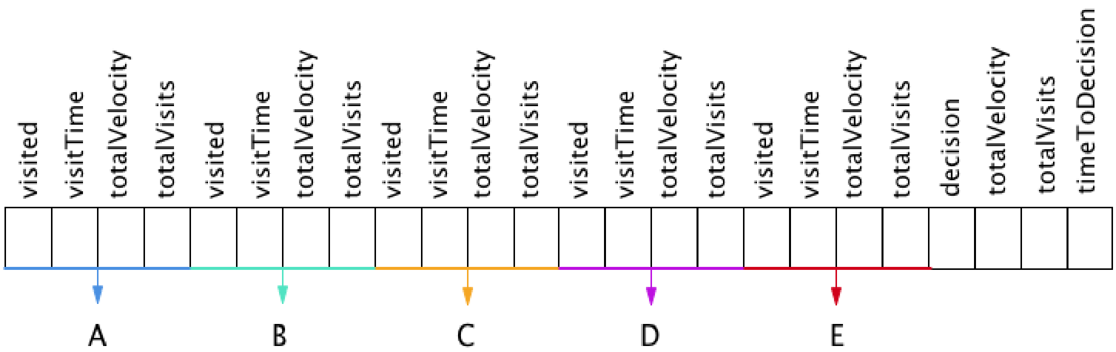



Figure 4.6: Game Telemetry Feature Array.

We compressed each array of mouse events in the *Telemetry Miner* to an array of desired game telemetry features with the *Telemetry Compressor*. Each array in the sample data above represents twenty-four game telemetry features (see Figure 4.6). The arrays include the following information:

- *visited<A, B, C, D, and E>*: A boolean value stating whether the area was visited.
- *totalTime<A, B, C, D, and E>*: The total time spent in the area in milliseconds.
- *totalVelocity<A, B, C, D, and E>*: The total mouse velocity in the area.
- *totalVisits<A, B, C, D, and E>*: The total visits to the area.
- *click*: The decision in a dialogue (i.e., A, B, C, D, or E for silence).

- *totalVelocity*: The total mouse velocity before the dialogue decision.
- *totalVisits*: The total visits to each area.
- *timeToDecision*: Total time before the mouse was moved to make a decision, in milliseconds.

### 4.3.2 EEG

We exported each player's EEG data from *eego* [64] and imported it into the EEG analytics software *ASA* [65]. We also imported a standard head model and electrode positions into *ASA* to prepare for analysing the data. In this step, it would have been ideal to have used the 3D head models acquired from each participant instead of the standard one, but unfortunately, a misunderstanding resulted in the acquired models being unusable.

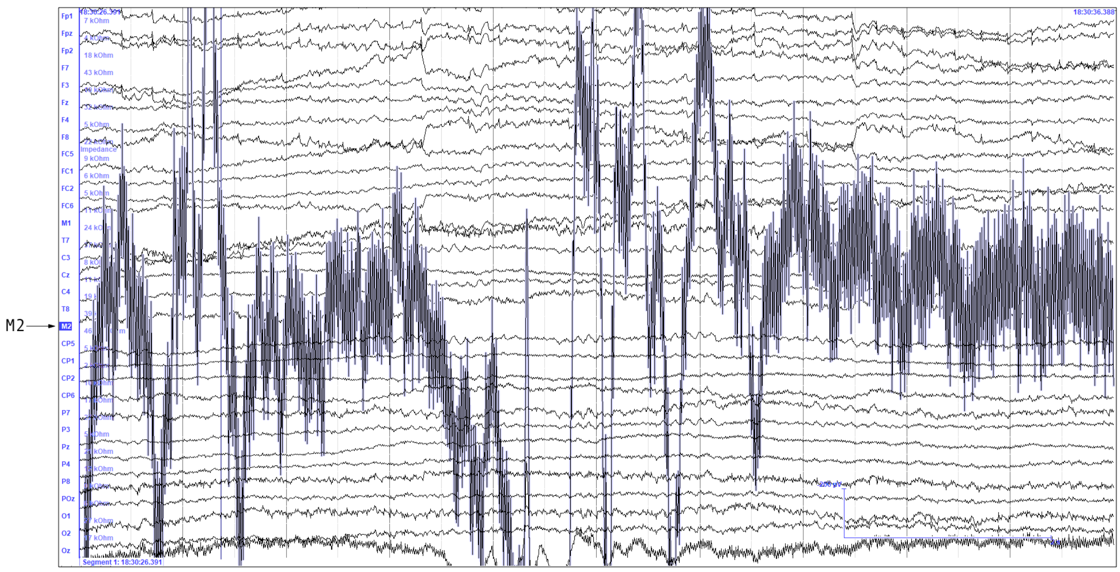



Figure 4.7: EEG recording displaying a contaminated electrode M2.

We started processing the EEG data by visually inspecting it for any obviously contaminated electrodes. In those cases, we removed the electrode from the data set before conducting further analysis (see Figure 4.7). We filtered the EEG data with a Band-Pass filter from 0.3 Hz to 45 Hz, which resulted in the removal of undesired frequencies. We also filtered the data with a Band-Stop Filter from 49 Hz to 51 Hz. This filter, also known as a notch filter, is used to remove noise caused by electrical equipment. Next, we detected artefacts by testing for amplitudes with upper and lower limits of -100 microvolts ($\mu$V) and +100 $\mu$V and rejected them from our data set. An example of eye blink artefacts identified in *ASA* can be seen in Figure 4.8.


Figure 4.8: EEG recording with eye blink artefacts (marked with a pink shade).

For each dialogue event in the game, we performed a Fast Fourier Transform (FFT) analysis. We used the 28 decision time stamps acquired from the *Telemetry Analyzer* and defined the corresponding time intervals containing five seconds up until a player chose a response in a dialogue event. The block length of the FFT was set to 1.024 s to keep the variables during processing to a minimum, resulting in quick measurement repetitions with a coarse frequency resolution. We rejected artefacts within the FFT and normalised the power spectra. Finally, we exported the FFT analysis for each event for the two bands of interest, alpha (7.5-12 Hz) and beta (12.5-30 Hz), and extracted emotion features (see sample data in Section 4.3.3). We based our emotion feature extraction approach on methods presented by Ramirez et al. [27] which were later extended by Blaiech et al. [12]. The EEG cap they used was the Emotiv EPOC cap [45]; it consists of 14 electrodes and two reference electrodes, and is based on the 10/20 electrode placement system. As our EEG cap is based on the 10/10 system and contains 32 electrodes instead of 14 [66], we mapped the electrodes Blaiech et al. used for feature extraction to our electrode layout. The electrodes of interest on the Emotiv cap were six electrodes in the frontal lobe: AF3, AF4, F3, F4, FC6, F8, and one location in the parietal lobe: P8 (see Figure 4.9).

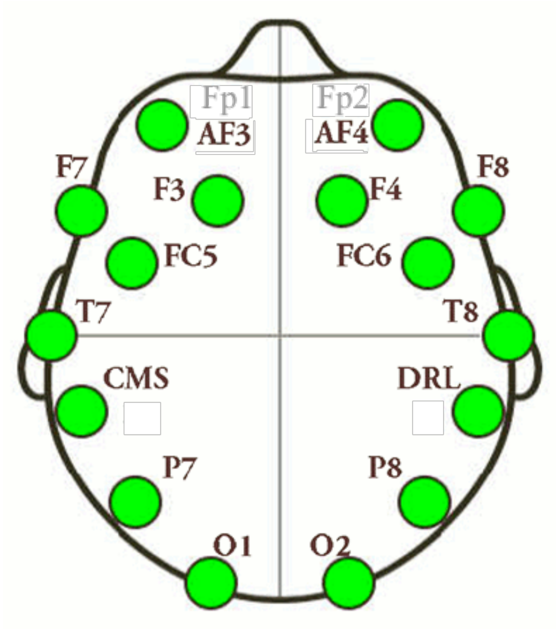



Figure 4.9: Electrode positions on the Emotiv EPOC cap. The grey text shows the location of the electrodes Fp1 and Fp2, which correspond to AF3 and AF4 in our study.

The 10/10 system is backwards compatible with the 10/20 system except for minor differences [67]. We mapped the electrodes AF3 and AF4 to the electrodes Fp1 and Fp2 as those electrodes were the closest to the positions of AF3 and AF4. To calculate arousal, valence, and dominance, we used the equations proposed by Ramirez et al. and Blaiech et al. [12], [27]. To calculate arousal, one calculates the ratio between beta ($\beta$) and alpha ($\alpha$) waves; it is identified by a high beta power as well as coherence in the parietal lobe, and a low alpha activity.

$$Arousal = \alpha(Fp1 + Fp2 + F3 + F4)/\beta(Fp1 + Fp2 + F3 + F4) \tag{4.1}$$

Valence is identified by comparing activation levels in the left prefrontal and the right prefrontal lobes. Inactivation in the left one indicates a negative emotion and inactivation in the right one indicates a positive emotion [43]. The inactivation is associated with a high alpha activity, because that is an indication of low brain activity and a decrease in beta waves.

$$Valence = \alpha F4/\beta F4 - \alpha F3/\beta F3 \tag{4.2}$$

Dominance is identified by an increase in the ratio between beta and alpha activity in the frontal lobe as well as an increase in beta activity in the parietal lobe.

$$Dominance = (\beta FC6/\alpha FC6) + (\beta F8/\alpha F8) + (\beta P8/\alpha P8) \tag{4.3}$$

To obtain the minimum value of arousal for each player, we performed the same calculation for arousal on the EEG baseline recordings of both eyes closed and eyes open [61]. Then, we normalised the values of arousal acquired from the baseline and the gameplay recording to values between 0 to 1 for each player. The baseline does not provide the minimum values of valence and dominance. Therefore, to normalise valence and dominance, we used the minimum values for each during the gameplay instead of the baseline. This normalisation allowed us to use values from multiple players, as the numeric values obtained from the FFT analysis differed greatly between each player. Sample calculations of arousal, valence, and dominance are presented in Section 4.3.3.

### 4.3.3 Sample of EEG Data

The following data sample shows results from an FFT analysis conducted on filtered EEG data for a five-second dialogue event. The fields in *ValuesTransposed* show FFT analysis for each electrode in the alpha and the beta band (marked as *0.000e+000* and *9.771e-001*). The highlighted values are the ones we use to calculate the three emotion features (arousal, valence, and dominance).

EEG Data - FFT Power Spectrum Normalised

```
# Time Frequency Data Export
SpectrumType  FFTPowerSpectrumNormalized
StartTime=  --:--:--.---
EndTime=  --:--:--.---
SamplingFrequency=  5.120000e+002
EpochLength=  1.024000e+000
NumberOfEpochs=  1
BandsName=  2 Bands (Alpha-Beta)
BandsDescription=
two bands 7.5 to 12.5 Hz (Alpha) and 12.5 to 30 Hz (Beta)
NumberOfBands= 2
BandsData
7.50  12.50
12.50  30.00

BandsNames
7.5 - 12.5 Hz
12.5 - 30.0 Hz
NumberOfChannels=  32
UnitMeas
ValuesTransposed
0.000e+000:  1.1007e-001  1.1404e-001  1.2539e-001  1.3815e-001
             5.3200e-002  5.2164e-002  6.5276e-002  1.3197e-001
             7.3727e-002  2.2893e-002  3.0865e-002  9.7599e-002
             1.3295e-001  9.7604e-002  3.0026e-002  7.8630e-003
             5.1545e-002  1.1612e-001  1.4499e-001  4.3226e-002
```

```
             4.5867e-003  5.0144e-003  7.8908e-002  6.4002e-002
             5.5455e-003  7.7527e-003  3.9665e-002  1.9489e-001
             4.9524e-002  8.2759e-002  4.6000e-002  5.3284e-002
9.771e-001:  7.4231e-002  8.1237e-002  9.2017e-002  6.3959e-002
             3.1600e-002  3.8681e-002  6.6716e-002  9.9810e-002
             3.1671e-002  1.7792e-002  2.2243e-002  8.7542e-002
             6.5330e-002  2.3061e-001  1.4624e-002  6.8285e-003
             2.0024e-002  6.1163e-002  7.3708e-002  3.9870e-002
             3.0001e-003  2.9987e-003  2.7836e-002  6.7414e-002
             3.3141e-002  3.8817e-003  1.5593e-002  4.2857e-002
             2.2614e-002  6.2334e-002  4.3809e-002  5.6550e-002
NumberOfDataSets=  1
DataSets
C:\path\to\data\eeg-data.cnt
Labels
Fp1  Fpz  Fp2  F7  F3  Fz  F4  F8  FC5  FC1  FC2  FC6
M1  T7  C3 Cz  C4 T8  M2  CP5  CP1  CP2  CP6  P7
P3   Pz  P4  P8  POz  O1  O2  Oz
NumberEpochsPerChannel
5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
NumberOffsetEpochsPerChannel
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
NumberDataSetsPerChannel
1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
```

---

From the data sample above, we can calculate arousal, valence and dominance with Equations 4.1, 4.2, and 4.3:

$$Arousal = \frac{0.110070+0.125390+0.053200+0.065276}{0.074231+0.092017+0.031600+0.066716} = 1.337809$$

$$Valence = \frac{0.065276}{0.066716} - \frac{0.053200}{0.031600} = -0.705128$$

$$Dominance = \frac{0.087542}{0.097599} + \frac{0.099810}{0.131970} + \frac{0.042857}{0.194890} = 1.873168$$

Values after applying normalisation with minimum and maximum values $\{0.281725; 2.874338\}$ for arousal, $\{1.449725; 1.058190\}$ for valence, and $\{1.317771; 9.962032\}$ for dominance:

$Arousal = 0.407343$

$Valence = 0.296899$

$Dominance = 0.064250$

These values thus represent the estimates of the player's emotion features during the dialogue event that we have been using as our example.

## 4.4 Emotion Prediction

We trained and tested three machine learning algorithms with the game telemetry data and the values of the emotion features (arousal, valence, and dominance). We aimed for the algorithm to predict the emotion features from the game telemetry. Due to artefact rejection in the EEG analysis, we had to remove the corresponding dialogue events from the player's data. As a result, we had 26.5 dialogue events per player on average instead of 28. We preprocessed the game telemetry data to account for the differences in the units, scaling our data between values of zero and one for each player separately. We categorised the values of arousal, valence, and dominance into three categories: 0, 0.5, and 1, to represent low, medium, and high values. We classified every feature value less than 0.33 as a 0, values equal to or greater than 0.33 and less than 0.66 as a 0.5, and every value equal to or greater than 0.66 as a 1. We'll refer to these categories as Low, Medium and High in Chapter 5.

Since our data set was fairly small, we decided to classify the data by using Decision Tree, Nearest Neighbour, and Naive Bayes classifiers. We trained and tested the classifiers in two ways. First, we randomly removed two players' data sets which each had twenty-eight dialogue events in total from the training data and trained on data from the remaining eighteen players, and tested the classifier on the two data sets we left out. We also trained and tested all of the twenty data sets via a 10-fold cross validation. We present our reasons for doing so in Section 5.2. To classify the data, we set up two different problems. We trained classifiers for each emotion feature separately, and we also trained a multi-target classifier for predicting the values of arousal, valence, and dominance in conjunction. As no other research has been conducted on a problem like ours, we strove to achieve better results than a uniform random classifier would produce. We present the results of our tests in Chapter 5.

## Summary

In this chapter, we explained and exemplified our methodology for predicting players' emotions from their game telemetry. We started by discussing why we chose to use the video game *The Wolf Among Us* in our user study. We gave a detailed description of our data acquisition protocol which consisted of obtaining data on demographics, EEG recording, mouse tracking, screen recording, a self-assessed emotion questionnaire, and a 3D head model. We presented our feature extraction methods for both the game telemetry and the emotion features from the EEG recording. Finally, we described how we used machine learning algorithms to predict emotion features from the game telemetry.

# Chapter 5

# Evaluation

We evaluated our emotion prediction method by comparing its results from the machine learning tasks to the results of a uniform random predictor. Due to artefact rejection in the EEG analysis, EEG data from some dialogue events was unavailable for some players, and we were forced to omit those events from our data set. This resulted in 26.5 dialogues per player on average instead of 28, and left us with 530 data samples in total.

## 5.1 Demographics

Twenty participants took part in our study. 13 participants were male and 7 were female, with a median age of 25-29 years (see Figure 5.1). Most of the participants were right handed, but 15% were left handed. Language processing for left handed people tends to take place in different brain hemispheres than for right handed people. Therefore, many EEG studies exclude left handed people from their studies [68]. Studies have also suggested that emotions may be measured differently for left handed people [69]. We tested our data for such effects by excluding the data from the three left handed people in our study and found no improvement in correctly classified instances. Therefore, we concluded that the data did not affect the prediction negatively and kept the data from the left handed people in our data set. Most of the participants (65%) said they spent one hour or more per day playing video games and the same percentage had not played a video game from *Telltale Games* before. Only one participant claimed to be an expert at games from *Telltale Games*. As stated before, none of the participants had played *The Wolf Among Us* before, since this was a prerequisite for taking part in the study. Another requirement was to be 18 years of age or older, since the video game's ESRB content rates the game as "Mature," stating that the content is suitable for ages 17 and up [70].

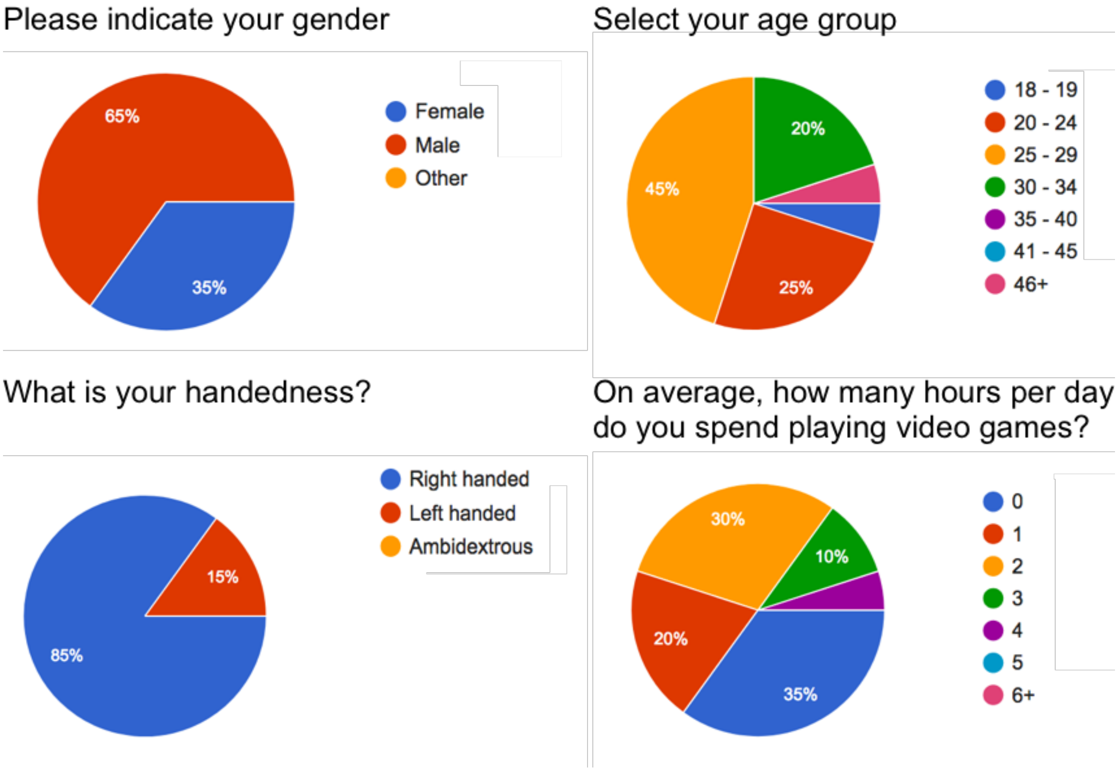



Figure 5.1: Statistics from the demographics questionnaire.

## 5.2 Emotion Prediction Results

For each emotion feature (arousal, valence, and dominance), we trained and tested three different machine learning algorithms: a pruned C4 Decision Tree [71], a Naive Bayes classifier, and a Nearest Neighbour classifier in the machine learning workbench *Weka* [72]. We also conducted a multi-target classification for predicting arousal, valence, and dominance at the same time with a Bagging MT classifier [73] in the machine learning toolkit *Meka* [74]. For all of the classification instances, we fed twenty-four game telemetry values as input to the algorithms and the emotion features as output, for all of the dialogue events that were recorded. We trained and tested the data in two different ways. Firstly, we partitioned the data samples into a training set from eighteen players to train the classifiers and a test set from two players for evaluation. To ensure that we would evaluate the classifiers' accuracy using every dialogue event in the game, we made sure that the data from players in the test set had twenty-eight dialogue events each. The other method we used was a repeated 10-fold cross-validation of the data. This method also partitions the data into a training set and a test set but does it randomly into ten equal samples. One sample is retained as test data, and the remaining samples are used as training data. This training and testing method is then repeated ten times such that each of the 10 partitions are used once as a test set. The results from the 10 classifiers are then averaged to produce a single estimation.

When training the Decision Tree, we pruned the tree with different values of confidence factors and a minimum number of instances per leaf [71]. The value of the confidence factor can be between 0.0 to 0.5. Lower values incur more pruning to the tree. The value of the minimum number of instances per leaf is used to determine how often the branches of the tree can be split into different leafs. The default settings for these values in *Weka* are 0.25 for the confidence factor and two is the minimum number of instances per leaf. We report the results of using these values to train the Decisions Tree in Tables 5.1 to 5.6. We also report the best results of changing these values to incur heavier or lighter pruning to the tree. The different values we used are reported in the tables as *C* for confidence factor and *M* for minimum number of instances per leaf.

A common rule of thumb for determining the number of neighbours to use in a Nearest Neighbour classifier is to use the square root of the total number of samples in the training set [75]. However, this rule does not always provide the highest accuracy and therefore we tried different numbers of neighbours to see which one gave the highest accuracy. In our training and test split of the data, we had 474 data samples in the training data. According to the rule of thumb, we should use 21 neighbours. We used this value as a reference to determine the optimal amount of neighbours which are listed in the results in Tables 5.1 to 5.6.

The data distribution for values of arousal, valence, and dominance is shown in Figures 5.2, 5.4 and 5.3. Most instances for arousal and dominance were Low. Due to this bias, a simple predictor could predict arousal as Low with 58.9% accuracy and dominance as Low with 65.3% accuracy. The values of valence were distributed more evenly between the classes with the highest percentage as Medium (40.0%). Therefore, a simple predictor would achieve similar results as a uniform random predictor when predicting valence.

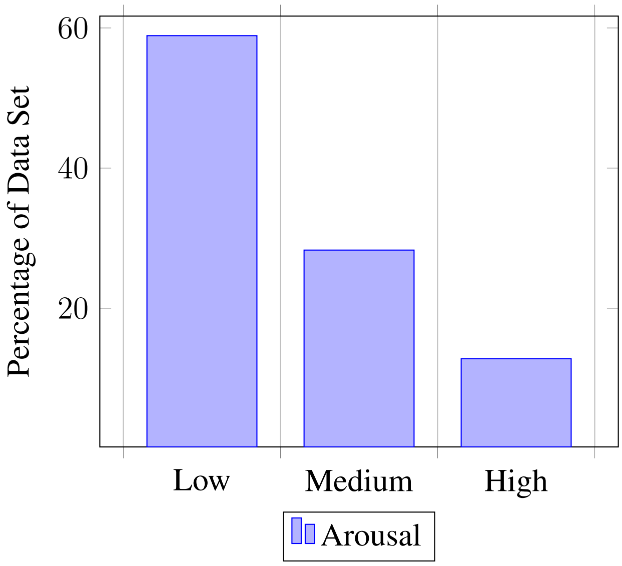



Figure 5.2: Distribution of arousal.

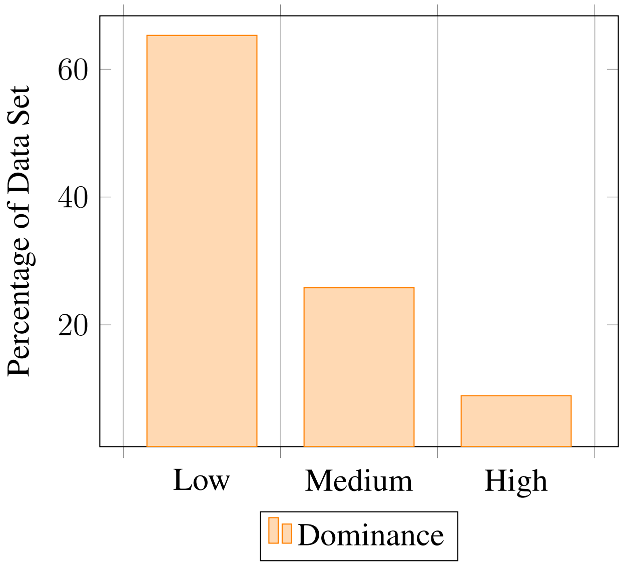



Figure 5.3: Distribution of dominance.

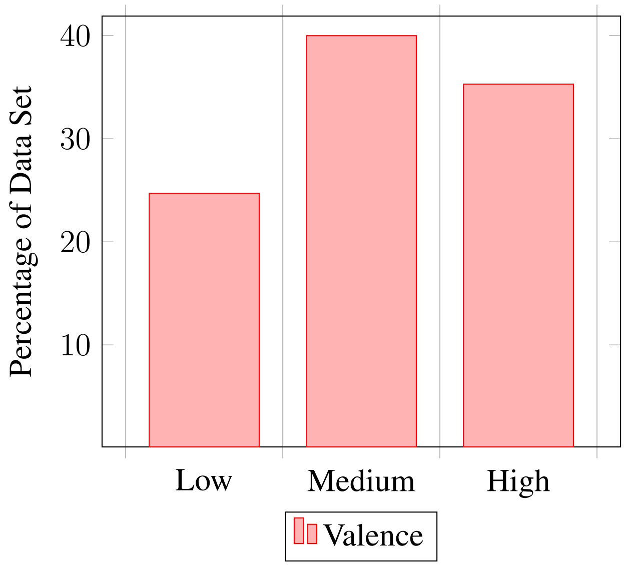



Figure 5.4: Distribution of valence.

The results from each classifier are presented in Tables 5.1 to 5.6. We determined the accuracy of each classifier by calculating the sum of correctly classified instances divided by the total number of classifications. We also calculated the weighted average precision and recall (or sensitivity) for each class (Low, Medium, and High). We obtained precision by dividing the number of true positives by the number of true positives and false positives [76]. Recall is the number of true positives divided by the number of true positives and the number of false negatives [76]. We did these calculations for each classifier except the multi-target classifier because Meka [74], the software we used to run the classification, did not report enough information for us to calculate the precision or recall.We compared our results with the results of a uniform random predictor, which, in our case of predicting three classes for each feature, would achieve 33.3% accuracy. We also compared them with the results that the simple predictors that we discussed previously would achieve.

The highest accuracy for predicting arousal with our custom training and testing split was with the Nearest Neighbour classifier of 21 neighbours (see Table 5.1). It correctly classified 75.0% of the instances of arousal and also reported high precision overall but a rather low recall for the class Medium. None of the classifiers achieved satisfactory results in classifying values as High.

| | Precision | | | Recall | | | |
|---|---|---|---|---|---|---|---|
| **Classifier** | **Low** | **Medium** | **High** | **Low** | **Medium** | **High** | **Accuracy** |
| Decision Tree C=0.25, M=2 | 75.9% | 16.7% | 0.0% | 53.7% | 65.5% | 0.0% | 44.6% |
| Decision Tree C=0.10, M=2 | 74.1% | 50.0% | 0.0% | 97.6% | 7.7% | 0.0% | 73.2% |
| Naive Bayes | 81.8% | 24.0% | 0.0% | 43.9% | 46.2% | 0.0% | 42.9% |
| Nearest Neighbour k=21 | 74.5% | 100.0% | 0.0% | 100.0% | 7.7% | 0.0% | 75.0% |
| Bagging MT | N/A | N/A | N/A | N/A | N/A | N/A | 57.1% |

Table 5.1: Results per class (Low, Medium, and High): Arousal - Training / Testing Split.

The highest accuracy for predicting arousal with the 10-fold cross-validation was with the Nearest Neighbour classifier of 29 neighbours (see Table 5.2). It correctly classified 59.1% of the instances of arousal. However, the Decision Tree with a confidence factor of 0.10 and a minimum number of two instances per leaf achieved a better performance overall. It achieved a lower accuracy of 58.1% but reported higher precision and recall overall.

| | Precision | | | Recall | | | |
|---|---|---|---|---|---|---|---|
| **Classifier** | **Low** | **Medium** | **High** | **Low** | **Medium** | **High** | **Accuracy** |
| Decision Tre C=0.25, M=2 | 59.6% | 30.9% | 17.4% | 68.9% | 25.3% | 11.8% | 49.2% |
| Decision Tree C=0.10, M=2 | 59.5% | 35.3% | 22.2% | 96.2% | 4.0% | 2.9% | 58.1% |
| Naive Bayes | 58.8% | 25.9% | 9.9% | 61.9% | 14.0% | 17.6% | 42.6% |
| Nearest Neighbour k=21 | 59.3% | 25.0% | 0.0% | 98.4% | 0.2% | 0.0% | 58.5% |
| Nearest Neighbour k=29 | 59.0% | 100.0% | 0.0% | 100.0% | 0.1% | 0.0% | 59.1% |
| Bagging MT | N/A | N/A | N/A | N/A | N/A | N/A | 48.3% |

Table 5.2: Results per class (Low, Medium, and High): Arousal - 10-fold Cross-validation.

As mentioned before, a simple predictor could predict arousal as Low with 58.9% accuracy. Our results with our custom training and testing split with the Nearest Neighbour classifier of 21 neighbours report a higher accuracy (75.0%) than both the results of a simple predictor and a uniform random predictor.

The classifiers did not perform as well when predicting valence (see Tables 5.3 and 5.4). The highest accuracy with our custom training and testing split was achieved with the Bagging multi-target classifier (58.9%) (see Table 5.3). Of the separate classification models, the Decision Tree with a confidence factor of 0.10 and a minimum number of two instances per leaf achieved the highest accuracy (53.6%) and overall precision and recall.

| | **Precision** | | | **Recall** | | | |
|---|---|---|---|---|---|---|---|
| **Classifier** | **Low** | **Medium** | **High** | **Low** | **Medium** | **High** | **Accuracy** |
| Decision Tree C=0.25, M=2 | 25.0% | 57.1% | 53.3% | 12.5% | 19.0% | 88.9% | 51.8% |
| Decision Tree C=0.10, M=2 | 50.0% | 55.6% | 53.3% | 12.5% | 23.8% | 88.9% | 53.6% |
| Naive Bayes | 16.7% | 37.5% | 33.3% | 25.0% | 57.1% | 14.8% | 32.1% |
| Nearest Neighbour k=21 | 8.3% | 43.5% | 42.9% | 12.5% | 47.6% | 33.3% | 35.7% |
| Bagging MT | N/A | N/A | N/A | N/A | N/A | N/A | 58.9% |

Table 5.3: Results per class (Low, Medium, and High): Valence - Training / Testing Split.

The highest accuracy with the 10-fold cross-validation for predicting valence was achieved with the Decision Tree with a confidence factor of 0.4 and a minimum number of two instances per leaf achieved (see Table 5.4). It achieved accuracy of 38.3% and a high percentage of precision and recall overall.

| | **Precision** | | | **Recall** | | | |
|---|---|---|---|---|---|---|---|
| **Classifier** | **Low** | **Medium** | **High** | **Low** | **Medium** | **High** | **Accuracy** |
| Decision Tree C=0.25, M=2 | 29.6% | 42.0% | 38.8% | 34.4% | 39.6% | 36.9% | 37.4% |
| Decision Tree C=0.40, M=2 | 30.5% | 43.3% | 38.5% | 32.8% | 43.9% | 35.8% | 38.3% |
| Naive Bayes | 27.7% | 45.5% | 35.3% | 76.3% | 21.7% | 12.8% | 32.1% |
| Nearest Neighbour k=21 | 21.8% | 39.7% | 39.0% | 16.8% | 48.1% | 35.8% | 36.0% |
| Bagging MT | N/A | N/A | N/A | N/A | N/A | N/A | 34.0% |

Table 5.4: Results per class (Low, Medium, and High): Valence - 10-fold Cross-validation.

Valence was more evenly distributed than arousal and therefore a simple predictor would achieve similar results as a uniform random predictor, or 40.0% accuracy. Our results with the custom training and testing split with the multi-target classifier (58.9%) show that we achieve higher accuracy than the uniform random predictor and the simple predictor.

The highest accuracy for predicting dominance with our custom training and testing split was achieved with the Bagging multi-target classifier (51.8%) (see Table 5.5). Of the separate classification models, the Decision Tree with a confidence factor of 0.20 and a minimum number of two instances per leaf achieved the highest accuracy (50.0%) and overall precision and recall.

| | **Precision** | | | **Recall** | | | |
|---|---|---|---|---|---|---|---|
| **Classifier** | **Low** | **Medium** | **High** | **Low** | **Medium** | **High** | **Accuracy** |
| Decision Tree C=0.25, M=2 | 49.1% | 0.0% | 0.0% | 96.4% | 0.0% | 0.0% | 48.2% |
| Decision Tree C=0.20, M=2 | 50.0% | 50.0% | 0.0% | 96.4% | 4.8% | 0.0% | 50.0% |
| Naive Bayes | 50.0% | 0.0% | 0.0% | 89.3% | 0.0% | 0.0% | 44.6% |
| Nearest Neighbour k=21 | 50.0% | 0.0% | 0.0% | 100.0% | 0.0% | 0.0% | 50.0% |
| Bagging MT | N/A | N/A | N/A | N/A | N/A | N/A | 51.8% |

Table 5.5: Results per class (Low, Medium, and High): Dominance - Training / Testing Split.

With the 10-fold cross-validation, the Nearest Neighbour classifier with 28 neighbours achieved the highest accuracy (65.5%) for predicting dominance (see Table 5.6). It also achieved the overall highest precision but a low recall for predicting values as Medium.

| | Precision | | | Recall | | | |
|---|---|---|---|---|---|---|---|
| **Classifier** | **Low** | **Medium** | **High** | **Low** | **Medium** | **High** | **Accuracy** |
| Decision Tree C=0.25, M=2 | 65.2% | 25.8% | 0.4% | 76.9% | 18.2% | 2.1% | 55.1% |
| Decision Tree C=0.20, M=2 | 65.4% | 18.9% | 7.1% | 87.6% | 7.3% | 2.1% | 59.2% |
| Naive Bayes | 65.7% | 30.1% | 9.5% | 18.8% | 22.6% | 66.0% | 24.0% |
| Nearest Neighbour k=21 | 65.1% | 16.7% | 0.0% | 98.6% | 1.0% | 0.0% | 64.5% |
| Nearest Neighbour k=28 | 65.4% | 100.0% | 0.0% | 100.0% | 1.0% | 0.0% | 65.5% |
| Bagging MT | N/A | N/A | N/A | N/A | N/A | N/A | 58.4% |

Table 5.6: Results per class (Low, Medium, and High): Dominance - 10-fold Cross-validation.

The two samples we provided to the test set may not have reflected a typical dominance response, which could explain why 10-fold cross-validation gave us better results when predicting dominance (see Table 5.6). Our results with the Nearest Neighbour classifier with 28 neighbours achieved higher accuracy (65.5%) than the accuracy of a uniform random predictor. However, a simple predictor could predict dominance as Low with 65.3% accuracy. Our results are only slightly better than the results of this simple predictor.

# Summary

In this chapter, we presented the work we conducted to evaluate our emotion prediction method. We discussed results from the demographics questionnaire and why we included data from left handed people in the study. We presented the results from several different classifiers on which we trained and tested our data, which showed that we were able to predict arousal, valence, and dominance with highest accuracies of 75.0%, 58.9%, and 65.5% respectively.

# Chapter 6

# Discussion

In this chapter, we discuss our proposed approach and its performance in four different sections. Firstly, we discuss the results from our emotion prediction method (Section 6.1). We then address the difficulties we faced in both our data acquisition phase (Section 6.2) and the data analysis phase (Section 6.3). We present the sources of inaccuracies in our data (Section 6.4) and finally, we present the limitations of our proposed approach and suggest ways to improve it in future work (Section 6.5).

## 6.1 Evaluation Results

The aim of our work was to formulate a method that could predict players' emotions from game telemetry with a higher accuracy than the results of a uniform random predictor. As our results state in Section 5.2, our best accuracies for predicting the emotion features were 75.0% for arousal, 58.9% for valence, and 65.5% for dominance. Therefore, we have achieved our goal of predicting emotional features with better results than a uniform random predictor and also the results of a simple predictor. By looking at the data distribution in Figures 5.2, 5.4, and 5.3, we can see a bias toward Low arousal and Low dominance. A simple predictor could classify arousal or dominance as Low with 58.9% accuracy for arousal and 65.3% for dominance. Taking this into consideration, our accuracy for arousal is still promising, but for predicting dominance, we only achieve slightly higher accuracy than the results of a simple predictor. The result for predicting arousal may suggest that the game telemetry we used offers a good indication of arousal levels (e.g., the player might move the mouse faster when arousal levels are high). However, further work is needed since the cross-validation accuracy for arousal was not as high. Our classifiers performed poorly when predicting dominance and our method of splitting the data with 10-fold cross-validation gave better results than our training and testing partitioning. The reasons for this may be that the two subjects we used as a testing set may have presented an atypical expression of dominance, leading to poorer classification results. Also, we might have achieved better results by using more data since the data we collected was biased towards Low dominance; data from twenty players may not be sufficiently representative of a wide range of players. Our classifiers had the lowest accuracy for predicting valence, although the result was satisfactory since it was well above 33.3% and our data showed little distribution bias.

We did not use the self-assessed emotion questionnaire in our prediction method or evaluation, due to a lack of foresight in the way it was set up. We only asked for emotions during certain scenes in the game, but each scene included several different dialogue events. Meanwhile, in our emotion prediction method, we predicted the emotions for each dialogue event separately. Since the result from each scene might have described a range of emotions that the player experienced across several dialogue events, we concluded that the self-assessed emotions would not give us reliable data for validating our methods. Furthermore, testing the calculations of arousal, valence, and dominance was beyond the scope of this study.

## 6.2 Challenges during Data Acquisition

To conduct the EEG recording in the data acquisition phase, we needed to be trained by a Technical Operator in The Icelandic Center for Neurophysiology at Reykjavik University. This technology is far from being "plug and play" and requires professional skills to record a high-quality EEG signal. For example, the EEG cap we used in the study was a wet electrode cap, meaning it required a certain amount of gel to be inserted in each electrode to reach the desired impedance for conductivity. This procedure can be very time-consuming, as all electrodes need to be filled correctly (32 in our case). Due to a misunderstanding, we recorded the EEG data with an incorrect electrode layout. As a result, we were unable to verify in *eego* the impedance of the electrodes we had filled with gel as the electrodes in the software showed a different placement. This slowed our process considerably, as sometimes it took extra time to find the electrodes in need of more gel. This issue could have been detected and resolved in the pilot study, but due to our inexperience, we only discovered it towards the end of the study. We ultimately sought help from *eego's* software support team to re-label our recorded EEG data to the correct electrode layout.

The EEG recording device limited us to 30 minutes of gameplay, as it was unable to record and store more data for each session. We do not believe this affected our results, since the player experienced a broad range of emotional content during the 30 minutes of our study.

In one of the play sessions, we did not press the recording button properly in *eego*. As a result, we did not acquire any EEG recording from the gameplay and had to replace the player and conduct a new recording. As presented in our Data Acquisition Protocol in Section 4.2.2, there were many things to consider in the study, and this incident motivated us to create a checklist such that mistakes like this would not happen again.

During one of the play sessions, we had to enter the laboratory because an inexperienced player had a hard time finishing a task in the game that required them to press the key Q repeatedly. We explained the controls better to the player, and they continued the game successfully. After this incident, we improved our explanation of the game controls and did not have an issue with this again.

Most players reported to us after the study that the EEG cap did not cause them any discomfort and that they forgot that they were wearing it after a few minutes of gameplay, which suggests that the cap did not seem to have affected the players' immersion or experience negatively, once the game was underway.

## 6.3 Challenges during Data Analysis

The analysis of the game telemetry was cumbersome. To analyse the data, we needed to develop four separate tools. The process would have been easier if we had been able to create a mod (i.e., an extension to a video game that alters some aspects of it) for *The Wolf Among Us*. By creating a mod, we could have developed an extension to log the player's decisions and time stamps during game play. Such a mod would have rendered our *Telemetry Analyzer* unnecessary, and we wouldn't have to had to visually verify the correctness of the time stamps from the *Telemetry Analyzer* by viewing the screen recordings from each play session. *Telltale Games*, however, forbids the development of mods for their video games. We were thus unable to implement a mod.

During the analysis of the EEG data, the user interface of *ASA* caused a significant slow down of the process. We were able to automate most of our processes with scripts, but it was not possible to automate exporting the FFT analysis that we conducted for the 530 dialogue events. Doing this step manually took a significant amount of time.

A common analysis method which we did not use in our EEG analysis is called *Independent Component Analysis* (ICA) and is used for eliminating artefacts [77]. The method decomposes the EEG recording into different components such that they can be inspected for artefacts. This requires a visual analysis by a trained specialist, which we did not have access to.

## 6.4 Sources of Inaccuracy

Due to several challenges in the study, the results might have been influenced by some inaccuracies. As we explained previously, the 3D head models taken from each player were unusable in the study. If we had been able to use them, they would have given us the exact positions of electrodes for each player, which would have likely given us more accurate results from the EEG analysis. To achieve the most accurate results, the impedance measured in each electrode should have been equal since differences in conductance makes some signals stronger.

The equations we used to estimate the values of the emotion features arousal, valence, and dominance were originally performed on an Emotiv EPOC cap that is based on a different electrode placement system from the WaveGuard cap we used. We had to map two of the electrodes that were used in the equations to electrodes that had a slightly different placement on our cap. As the electrode placement was not completely compatible, this might have introduced some inaccuracy to our results.

One of the game telemetry features that we acquired with the *Telemetry Compressor* was the total visits to each area during a dialogue event. This feature also included visits to the area outside of the response buttons (E). It might have given us more concise information to exclude the total visits to the outside area (E), since this area does not necessarily tell us if the player was considering multiple responses or not.

## 6.5 Limitations and Future Work

The main limitation of our method is that acquiring and processing EEG data is both time-consuming and labour-intensive. As such, it was only feasible for us to gather data from twenty players. This issue could be addressed in future work by using a larger team to obtain more data. Preferably, the study would be conducted with a dry electrode EEG cap to save time on gel insertions and impedance measures. Our data was biased toward Low for arousal and dominance. Future work might address this issue by not only including a large data set but also by normalising the distribution of the data before splitting it into categories. As stated in Section 5.2, we did not exclude left handed players from the data sets. Left handed players were only 15% of our sample, and our tests showed that they did not affect the results in a significant way. One of the left handed people was also the oldest of our players, and when we only excluded him, we achieved better results. Future work might limit the players' ages to a particular range to achieve better results.

To account for possible inaccuracies in this study, conducting it again with correct head models might improve our results for predicting the emotion features (arousal, valence, and dominance). We decided to not use the results from the self-assessed emotion questionnaire in our emotion prediction method. We might have been able to use them if we had shaped the questionnaire to have it ask about the player's emotional state for every dialogue event instead of certain scenes in the game. By including the results in our data set, we might have achieved higher prediction accuracies.

In this study, we only measured physiological response acquired with EEG. To achieve a better understanding of the player's emotional state, we might add more physiological sensors such as Heart rate variability (HRV), Galvanic Skin Response

(GSR) or EMG. McCraty et al. showed that positive emotions motivate alterations in HRV [24]. HRV could therefore be used to strengthen our calculations of valence. GSR measures arousal more directly, via skin conductance changes due to the amount of sweating [22]. Adding EMG measurements might have also improved our results since other studies have reported a high accuracy in classifying emotional features measured with EMG [78], [79].

The game telemetry features used in this study may be revised and extended. For example, it might strengthen the emotion prediction to classify each game event as being favoured by a particular player type or a personality (e.g., violent).

Our results show that it is possible to predict a player's emotional features from their game telemetry more accurately than a uniform random predictor. Incorporating our method in an actual player model might give us a better idea of the feasibility of our method in the context of player modelling. Once improved, our method could also be utilised in user studies before releasing video games, to test the emotional experiences that they evoke. In that way, the video game creators could understand better which aspects of their video game create which emotions within their players and use this information to improve players' experience.

# Chapter 7

# Conclusion

In this dissertation, we explored the interactive narrative aspect of video games and how interactive systems use player modelling in various ways to adjust a storyline to suit a player's type, preferences, or personality. Player modelling techniques often take the player's in-game actions as an input, but the player's emotions may also provide useful information. The player's in-game actions are highly influenced by their *feelings*, and how they feel is the player's interpretation of their *emotions* (i.e., their physiological response). People are often not very good at describing their emotions and, therefore, to evaluate a player's emotions, we proposed to predict their physiological response from their in-game actions.

Our proposed methodology predicts a player's emotions from their game telemetry (i.e., their in-game actions during a narrative decision). Our method consists of acquiring a player's game telemetry by tracking their mouse events and estimating their emotions by measuring their brain's electrical activity with EEG. Next, several machine learning algorithms can be trained and tested with the game telemetry as input and emotion features (arousal, valence, and dominance) as output.

To test our methodology, we conducted a study with twenty participants where they played the video game *The Wolf Among Us* by *Telltale Games*. During gameplay, we tracked their mouse and keyboard events and recorded their brain's electrical activity with an EEG cap.

The main limitation of our method was the time-consuming process of acquiring and analysing EEG data. It remains challenging to gather a large enough data set to represent a wide range of typical game telemetry and the corresponding emotions. Future work might address this issue by using a larger research team and by using a dry electrode EEG cap to save time on gel insertions.

We trained and tested three algorithms to predict each emotion feature separately from game telemetry, but also together with a multi-target classifier. We achieved the highest performance by predicting each feature separately with 75.0% accuracy of predicting arousal (custom training/testing split) and 65.5% for dominance (10-fold cross-validation). For predicting valence, we achieved the highest result of 58.9% (custom training/testing split) with a multi-target classifier. Our results showed that for at least one way of splitting our training and testing data, our method could predict players' emotion features from their game telemetry with a higher accuracy than both the results of a uniform random predictor and a simple predictor.

# Bibliography


[1] J. Gottschall, *The Storytelling Animal: How Stories Make Us Human*. Houghton Mifflin Harcourt, 2012.

[2] H. Bolin, “The Re-Generation of Mythical Messages: Rock Art and Storytelling in Northern Fennoscandia”, *Fennoscandia archaeologica*, vol. XXVII, pp. 21–34, 2010.

[3] S. Kraft, “He Chose His Own Adventure”, *The Day*, p. 6, 1981.

[4] H. Yu and M. Riedl, “Data-Driven Personalized Drama Management”, in *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, The AAAI Press, 2013, pp. 191–197.

[5] D. Thue, V. Bulitko, M. Spetch, and E. Wasylishen, “Interactive Storytelling: A Player Modelling Approach”, in *Proceedings of the Third AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, ser. AIIDE’07, AAAI Press, 2007, pp. 43–48.

[6] M. S. El-Nasr, “Interaction, Narrative, and Drama: Creating an Adaptive Interactive Narrative using Performance Arts Theories”, *Interaction Studies*, vol. 8, no. 2, pp. 209–240, 2007.

[7] Telltale Games, *Telltale Series*. [Online]. Available: `https://telltale.com/series/`.

[8] D. D. Franks, “The Neuroscience of Emotions”, in *Handbook of the Sociology of Emotions*, J. E. Stets and J. H. Turner, Eds. Boston, MA: Springer US, 2006, pp. 38–62, ISBN: 978-0-387-30715-2. DOI: `10.1007/978-0-387-30715-2_3`. [Online]. Available: `https://doi.org/10.1007/978-0-387-30715-2_3`.

[9] G. N. Yannakakis and J. Togelius, “Experience-Driven Procedural Content Generation”, *IEEE Transactions on Affective Computing*, vol. 2, no. 3, pp. 147–161, 2011.

[10] J. M. Kivikangas, G. Chanel, B. Cowley, I. Ekman, M. Salminen, S. Järvelä, and N. Ravaja, “A review of the use of psychophysiological methods in game research”, *Journal of Gaming & Virtual Worlds*, vol. 3, no. 3, pp. 181–199, 2011.

[11] J. Rodríguez, “EEG Emotion Recognition in Videogame Play”, Masters Thesis, Aalborg University, 2015.

[12] H. Blaiech, M. Neji, A. Wali, and A. M. Alimi, “Emotion recognition by analysis of EEG signals”, in *13th International Conference on Hybrid Intelligent Systems (HIS 2013)*, IEEE, Dec. 2013, pp. 312–318. DOI: 10.1109/HIS.2013.6920451.

[13] Atari, *Pong*, Video Game, Nov. 1972.

[14] W. Crowther, *Colossal Cave Adventure*, Video Game, 1976.

[15] C. Crawford, *Dragon Speech*, 1992. [Online]. Available: https://n42c.com/2014/03/02/thedragonspeech.

[16] M. Mateas and A. Stern, “Façade: An Experiment in Building a Fully-Realized Interactive Drama”, in *Game Developers Conference (GDC’03)*, vol. 2, 2003.

[17] M. Si, S. C. Marsella, and D. V. Pynadath, “Thespian: Using Multi-agent Fitting to Craft Interactive Drama”, in *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS ’05, The Netherlands: ACM, 2005, pp. 21–28.

[18] M. W. Fendt, B. Harrison, S. G. Ware, R. E. Cardona-Rivera, and D. L. Roberts, “Achieving the Illusion of Agency”, in *Interactive Storytelling: 5th International Conference, ICIDS 2012, San Sebastián, Spain, November 12-15, 2012. Proceedings*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 114–125, ISBN: 978-3-642-34851-8. DOI: 10.1007/978-3-642-34851-8_11. [Online]. Available: https://doi.org/10.1007/978-3-642-34851-8_11.

[19] Rockstar Games, *L.A. Noire*, Video Game, May 2011.

[20] Telltale Games, *The Walking Dead*, Video Game, Apr. 2012.

[21] T. L. Taylor, *Play Between Worlds: Exploring Online Game Culture*. The MIT Press, 2006.

[22] P. Lang, “The Emotion Probe”, *American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.

[23] R. W. Picard, *Affective Computing*. The MIT Press, 2000, pp. 161–162.

[24] R. McCraty, M. Atkinson, W. A. Tiller, G. Rein, and A. D. Watkins, “The effects of emotions on short-term power spectrum analysis of heart rate variability”, *The American Journal of Cardiology*, vol. 76, no. 14, pp. 1089–1093, 1995, ISSN: 0002-9149. DOI: http://dx.doi.org/10.1016/S0002-9149(99)80309-9. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0002914999803099.

[25] J. Gruber, D. S. Mennin, A. Fields, A. Purcell, and G. Murray, "Heart rate variability as a potential indicator of positive valence system disturbance: A proof of concept investigation", *International Journal of Psychophysiology*, vol. 98, no. 2, Part 2, pp. 240–248, 2015, Psychophysiological Science and the Research Domain Criteria, ISSN: 0167-8760. DOI: `https://doi.org/10.1016/j.ijpsycho.2015.08.005`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0167876015300210`.

[26] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, Engagement and Anxiety As Indicators for Adaptation to Difficulty in Games", in *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era*, ser. MindTrek '08, ACM, New York, NY, USA, 2008, pp. 13–17.

[27] R. Ramirez and Z. Vamvakousis, "Detecting Emotion from EEG Signals Using the Emotive Epoc Device", in *Brain Informatics: International Conference, BI 2012, Macau, China, December 4-7, 2012. Proceedings*, Springer Berlin Heidelberg, 2012, pp. 175–184. DOI: `10.1007/978-3-642-35139-6_17`. [Online]. Available: `https://doi.org/10.1007/978-3-642-35139-6_17`.

[28] B. Cowley, M. Filetti, K. Lukander, J. Torniainen, A. Henelius, L. Ahonen, O. Barral, I. Kosunen, T. Valtonen, M. Huotilainen, N. Ravaja, and G. Jacucci, "The Psychophysiology Primer: A Guide to Methods and a Broad Review with a Focus on Human–Computer Interaction", *Foundations and Trends® in Human–Computer Interaction*, vol. 9, no. 3-4, pp. 151–308, 2016.

[29] M. Teplan, "Fundamentals of EEG measurement", *Measurement Science Review*, vol. 2, no. 2, pp. 1–11, 2002.

[30] L. Ahonen and B. U. Cowley, "A short review and primer on electroencephalography in human computer interaction applications", *CoRR*, vol. abs/1609.00183, 2016. [Online]. Available: `http://arxiv.org/abs/1609.00183`.

[31] R. Wang, G. Blackburn, and M. Desai, "Accuracy of wrist-worn heart rate monitors", *JAMA Cardiology*, vol. 2, no. 1, pp. 104–106, 2017.

[32] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions", *J. Cognition & Emotion*, vol. 15, no. 4, pp. 487–500, 2001.

[33] T. Baumgartner, M. Esslen, and L. Jäncke, "From emotion perception to emotion experience: Emotions evoked by pictures and classical music", *International Journal of Psychophysiology*, vol. 60, no. 1, pp. 34–43, 2006. DOI: `https://doi.org/10.1016/j.ijpsycho.2005.04.007`.

[34] B. Reuderink, C. Muhl, and M. Poel, "Valence, arousal and dominance in the EEG during game play", *International journal of autonomous and adaptive communications systems*, vol. 6, no. 1, pp. 45–62, 2013, eemcs-eprint-23591, ISSN: 1754-8632. DOI: 10.1504/IJAACS.2013.050691.

[35] J. Kumar and J. Kumar, "Affective Modelling of Users in HCI Using EEG", *Procedia Computer Science*, vol. 84, pp. 107–114, 2016, Proceeding of the Seventh International Conference on Intelligent Human Computer Interaction (IHCI 2015), ISSN: 1877-0509. DOI: http://dx.doi.org/10.1016/j.procs.2016.04.073. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050916300886.

[36] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub., 1997.

[37] R. Davidson, P. Ekman, C. Saron, J. Senulis, and W. Friesen, "Approach-Withdrawal and Cerebral Asymmetry: Emotional Expression and Brain Physiology I", *Journal of Personality and Social Psychology*, vol. 58, no. 2, pp. 330–341, 1990.

[38] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music", *Psychophysiology*, vol. 44, no. 2, pp. 293–304, 2007. DOI: 10.1111/j.1469-8986.2007.00497.x.

[39] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Palo Alto, Ca.: Consulting Psychologists Press, 1978.

[40] R. J. Davidson and K. Hugdahl, "Baseline asymmetries in brain electrical activity predict dichotic listening performance", *Neuropsychology*, vol. 10, no. 2, pp. 241–246, 1996.

[41] N. Thammasan, K. Moriyama, K.-i. Fukui, and M. Numao, "Familiarity effects in EEG-based emotion recognition", *Brain Informatics*, vol. 4, no. 1, pp. 39–50, 2017. DOI: 10.1007/s40708-016-0051-5.

[42] J. A. Russel, "A circumplex model of affect", *J. Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[43] R. Davidson, "Anterior cerebral asymmetry and the nature of emotion", *Brain & Cognition*, vol. 20, no. 1, pp. 121–151, 1992.

[44] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-Time EEG-Based Emotion Recognition and Its Applications", *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artifical Intelligence and Lecture Notes in Bioonformatics)*, vol. 6670, pp. 256–277, 2011.

[45] EMOTIV Inc, *EMOTIV Epoc+*.

[46] D. Schwarz, S. Vivek, Z. Katie, and A. Christine, “Educational Neurogaming: EEG-Controlled Videogames As Interactive Teaching Tools For Introductory Neuroscience”, in *Experimental Artificial Intelligence in Games: Papers from the AI-IDE Workshop*, Center for Neuroengineering, 2014, pp. 49–52.

[47] Q. Wang, O. Sourina, and M. K. Nguyen, “EEG-based "Serious" Games Design for Medical Applications”, in *International Conference on Cyberworlds*, 2010, pp. 270–276. DOI: `10.1109/CW.2010.56`.

[48] D. Chen, J. James, F. S. Bao, C. Ling, and T. Fan, “Relationship Between Video Game Events and Player Emotion Based on EEG”, in *International Conference on Human-Computer Interaction*, Springer, 2016, pp. 377–384.

[49] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, “Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty”, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 6, pp. 1052–1063, 2011. DOI: `10.1109/TSMCA.2011.2116000`.

[50] L. Derbali and C. Frasson, “Prediction of Players Motivational States Using Electrophysiological Measures during Serious Game Play”, *2010 IEEE 10th International Conference on Advanced Learning Technologies (ICALT 2010)*, pp. 498–502, 2010.

[51] Telltale Games, *The Wolf Among Us*, Video Game, Nov. 2014.

[52] P. Theissen, *Mouse’n’Key*, Software, 2017. [Online]. Available: `https://github.com/pablotheissen/Mouse-n-Key`.

[53] *The Wolf Among Us - Customer Reviews*. [Online]. Available: `http://store.steampowered.com/app/250320/The%5C_Wolf%5C_Among%5C_Us/%5C#app%5C_reviews%5C_hash`.

[54] American Clinical Neurophysiology Society. (2006). Guideline 5: Guidelines for Standard Electrode Position Nomenclature, [Online]. Available: `https://www.acns.org/pdf/guidelines/Guideline-5.pdf`.

[55] eemagine Medical Imaging Solutions GmbH, *Data Sheet*, 2nd ed., Gubener Str. 47, 10243, Berlin, Germany, Dec. 2014.

[56] G. A. Light, L. E. Williams, F. Minow, J. Sprock, A. Rissling, R. Sharp, N. R. Swerdlow, and D. L. Braff, “Electroencephalography (EEG) and Event-Related Potentials (ERP’s) with Human Participants”, *Current protocols in neuroscience / editorial board, Jacqueline N. Crawley ... [et al.]*, vol. CHAPTER, no. Unit-6.2524, Jul. 2010. DOI: `10.1002/0471142301.ns0625s52`.

[57] Hugh Bailey, *Open Broadcaster Software*, Software, OBS Project.

[58] D. Watson, L. A. Clark, and A. Tellegen, “Development and validation of brief measures of positive and negative affect: The PANAS Scales”, in *Journal of Personality and Social Psychology*, vol. 54, 1988, pp. 1063–1070.

[59] P. Parente, *pyhook*, Software, 2008. [Online]. Available: `http://pyhook.sourceforge.net/doc_1.5.0/`.

[60] I. Ó. Jónsdóttir, *Telemetry Tracker*, Software, 2017. [Online]. Available: `https://github.com/Ingibjorg/TelemetryTracker`.

[61] R. J. Barry, A. R. Clarke, S. J. Johnstone, C. A. Magee, and J. A. Rushby, “EEG differences between eyes-closed and eyes-open resting conditions”, *Clinical Neurophysiology*, vol. 118, no. 12, pp. 2765–2773, 2007. DOI: `http://dx.doi.org/10.1016/j.clinph.2007.07.028`.

[62] ANT Neuro, *Xensor*, Software.

[63] R. W. Levenson, “Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity”, *Social Psychophysiology and Emotion: Theory and Clinical Applications*, pp. 17–42, 1988.

[64] ANT Neuro, *eego*, Software.

[65] ——, *ASA*, Software.

[66] ——, *Electrode Layouts*, Jun. 2017. [Online]. Available: `http://www.ant-neuro.com/products/waveguard/electrode-layouts`.

[67] V. Jurcak, D. Tsuzuki, and I. Dan, “10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems”, *NeuroImage*, vol. 34, no. 4, pp. 1600–1611, 2007, ISSN: 1053-8119. DOI: `http://dx.doi.org/10.1016/j.neuroimage.2006.09.024`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S1053811906009724`.

[68] R. M. Willems, L. V. der Haegen, S. E. Fisher, and C. Francks, “On the other hand: including left-handers in cognitive neuroscience and neurogenetics”, *Nature Reviews Neuroscience*, vol. 15, no. 3, pp. 193–201, Mar. 2014. [Online]. Available: `http://dx.doi.org/10.1038/nrn3679`.

[69] G. Brookshire and D. Casasanto, “Motivation and Motor Control: Hemispheric Specialization for Approach Motivation Reverses with Handedness”, *PLoS ONE*, vol. 7, no. 4, pp. 1–5, Apr. 2012. DOI: `10.1371/journal.pone.0036036`. [Online]. Available: `https://doi.org/10.1371/journal.pone.0036036`.

[70] ESRB, *The Wolf Among Us - ESRB Rating*. [Online]. Available: `http://www.esrb.org/ratings/Synopsis.aspx?Certificate=33709%5C&Title=The+Wolf+Among+Us`.

[71] J. Quinlan, *C4.5: Programs for Machine Learning*, ser. Morgan Kaufmann Series in Machine Learning. Morgan Kaufmann Publishers, 1993. [Online]. Available: `https://books.google.is/books?id=HExncpjbYroC`.

[72] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA Workbench", *Data Mining: Practical Machine Learning Tools and Techniques*, 2016.

[73] J. H. Zaragoza, L. E. Sucar, E. F. Morales, C. Bielza, and P. Larrañaga, "Bayesian Chain Classifiers for Multidimensional Classification", in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, ser. IJCAI'11, Barcelona, Catalonia, Spain: AAAI Press, 2011, pp. 2192–2197.

[74] J. Read, P. Reutemann, B. Pfahringer, and G. Holmes, "MEKA: A Multi-label/Multi-target Extension to Weka", *Journal of Machine Learning Research*, vol. 17, no. 21, pp. 1–5, 2016.

[75] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag New York, Inc., 1996, vol. 31.

[76] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed. Morgan Kaufmann, 2006, p. 361.

[77] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent Component Analysis of Electroencephalographic Data", in *Advances in Neural Information Processing Systems*, vol. 8, MIT Press, 1996, pp. 145–151.

[78] S. R. Vrana, "The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG", *Psychophysiology*, vol. 30, no. 3, pp. 279–286, 1993.

[79] A. Konar and A. Chakraborty, "Frequency Band Localization on Multiple Physiological Signals for Human Emotion Classification Using DWT", *Emotion Recognition: A Pattern Analysis Approach*, 2015.

# Appendix A

# Questionnaires

## A.1 Demographics

# Demographics Questionnaire

* Required

1. **Please indicate your gender** *
   *Mark only one oval.*
   - Female
   - Male
   - Other

2. **Select your age group** *
   *Mark only one oval.*
   - 18 - 19
   - 20 - 24
   - 25 - 29
   - 30 - 34
   - 35 - 40
   - 41 - 45
   - 46+
   - Other: ______________________

3. **What is your handedness?** *
   *Mark only one oval.*
   - Right handed
   - Left handed
   - Ambidextrous

4. **For how many years have you played video games?** *
   *Mark only one oval.*
   - 0
   - 0 - 5
   - 5 - 10
   - 10 - 15
   - 15 -20
   - 20 - 25
   - 26+
   - Other: ______________________

5. **On average, how many hours per day do you spend playing video games?** *
*Mark only one oval.*

- 0
- 1
- 2
- 3
- 4
- 5
- 6+
- Other: ____________

6. **I enjoy playing adventure games** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | | | | | | Strongly Agree |

7. **I enjoy playing action games** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | | | | | | Strongly Agree |

8. **Have you played any games from Telltale Games?** *
*Mark only one oval.*

- Yes
- No

9. **If yes, please rate your expertise at playing Telltale Games**
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Beginner | | | | | | Expert |

## A.2 Self-assessed emotions

The twenty questions below the first scene in the questionnaire were also applied to the other scenes.

# Wolf Among Us Questionnaire

For each of the following scenes, please indicate the extent to which you felt each of the listed emotions.

* Required

## The interaction with Toad


1. **Interested** *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | ◯ | ◯ | ◯ | ◯ | ◯ | Extremely |

2. **Distressed** *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | ◯ | ◯ | ◯ | ◯ | ◯ | Extremely |

3. **Excited** *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | ◯ | ◯ | ◯ | ◯ | ◯ | Extremely |

4. **Upset** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

5. **Strong** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

6. **Guilty** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

7. **Scared** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

8. **Hostile** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

9. **Enthusiastic** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

10. **Proud** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

11. **Irritable** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

12. **Alert** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

13. **Ashamed** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

14. **Inspired** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

15. **Nervous** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

16. **Determined** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

17. **Attentive** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | | | | | | Extremely |

18. **Jittery** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | Extremely |

19. **Active** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | Extremely |

20. **Afraid** *
*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Very Slightly or Not at All | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | Extremely |

## The fight with Woodsman

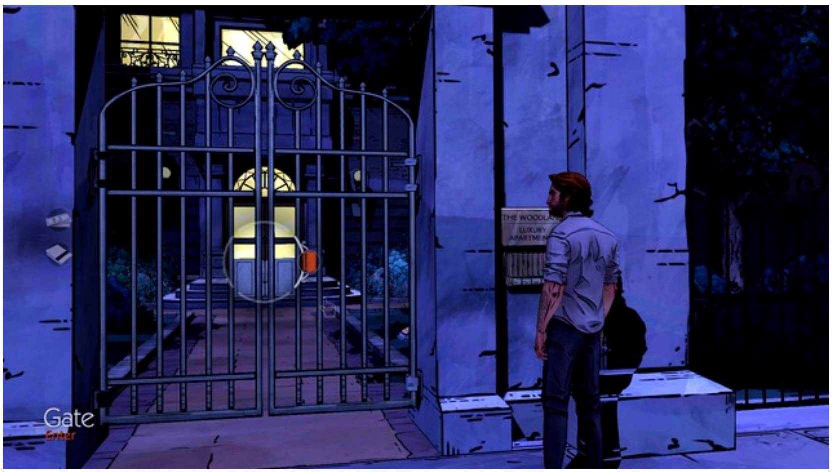

**...**

## The interaction with Faith

...

---

## Arriving home




...

---

## The interaction with Colin

...

## Finding Faith’s head

...

Powered by
Google Forms

School of Computer Science
Reykjavík University
Menntavegur 1
101 Reykjavík, Iceland
Tel. +354 599 6200
Fax +354 599 6201
www.ru.is