



Non-intrusive Measurement of Player Engagement and Emotions - Real-Time Deep Neural Network Analysis of Facial Expressions During Game Play

Dines Rae Selvig¹(✉)  and Henrik Schoenau-Fog²(✉) 

¹ Selvig Consultancy, Copenhagen, Denmark
dines@dinesselvig.com

² Aalborg University, Copenhagen, Denmark
hsf@create.aau.dk

Abstract. Prior research suggests and reveals that there is a correlation between human emotional responses and the subjective qualities of digital interactive experiences. Using facial analysis done by deep neural networks presents a true non-intrusive way of measuring emotional responses and engagement assessed as the desire to continue playing. This paper proposes a tool to measure emotional responses across eight different emotions and in real time of any game. The emotional recognition system achieves an accuracy of 98% and the continuation desire system achieves 93.3% accuracy in a pilot test with a two player game and 78.5% accuracy in a single player game. This forms a strong tool that shows a correlation between emotions and the continuation desire of a player, which can be used to evaluate engagement in games and digital interactive experiences, e.g. in critical stages of development of said content.

Keywords: Continuation desire · Conation · Emotion · Machine learning · Facial expressions · Design tools/technologies · Game and flow · Game immersion · Player engagement assessment

1 Introduction

The Game industry is a billion dollar industry, with video game revenue predicted to exceed over 180 billion US dollars by 2021 [60] and it is an industry, where content quality often dictates the revenue. It is estimated that with an improved infrastructure for testing and removal of bugs or low quality scenes, millions of dollars can be saved [44]. The QA (Quality Insurance) department needs more tools at hands for testing, not only for game mechanic bugs that makes a game unplayable but also for narrative bugs, which is an increasingly

Supported by Samsung Media Innovation Lab for Education (SMILE Lab).

part of every game. In the current research, a vast majority of interactive experience testing revolves around using interviews and observation of play testers to evaluate games and interactive experiences in terms of fun and engagement. Few attempts have been made to develop a non-intrusive framework for a more reliable testing using different data gathering techniques. Such data could give a deeper understanding of what a user experience during a test session. Many techniques rely on heavy equipment and complicated setups, which is not always applicable for smaller studios to use, or it is complicated to replicate a specific scientific method from an academic paper. As the rise of Machine Learning and computational power, it is now more accessible to use algorithms to find correlations in high dimensional data signals and give a better understanding of what these signals mean. Other than for testing purposes, such data could also be used for adaptive experiences in a real-time setting, such as changing the narrative or character behaviour or even lighting mood inside an interactive experience, to create a more tailored experience to the user.

When evaluating interactive experiences, one can measure, flow [10, 12], presence [31], immersion [29] or enjoyment [27], but one method of measure specifically developed for evaluating engagement in interactive experiences is the concept of continuation desire [45–47, 50]. Continuation desire is the desire to continue playing/watching digital content, which evaluate the real metric for content to be engaging and popular, namely retention [15, 58].

Emotion recognition is the process to identify an affective state of a person, and is getting increasing attention from researchers in human-computer interaction [52]. Such as the use for gaming experience evaluation [25], mental diagnosis [57], driving safety [56]. There is a vast amount of different data measures that can be used for emotion recognition such as audio, video and physiological measures. Using audio as emotion recognition have been used by researchers in different works, using spectral analysis as features to predict emotions in speech [7]. In terms of visual based recognition, different approaches have been explored such as older systems using binary patterns [34, 53], and image processing techniques to track landmark points [55] and newer methods such as convolutional neural networks [6, 21]. Typically individual researchers achieve high accuracies, in the area of 90–95%, within their own testings, but it is usually a result of sparse data, and commonly systems learn specific individuals which is used in the experiment, rather than a generalised solution [28, 30, 38, 61].

In this paper, the scope and focus is to develop such a tool to measure emotions and continuation desire both during development of interactive experiences and finished productions such as games. The measured data from a test session is then visualised as graphs to pinpoint emotional spikes and changes in continuation desire, these points can show elements in the interactive experience which spark these spikes.

The tool is equally relevant and applicable in adaptive game balancing or interactive experiences, as an input to control certain aspects of the environment such as the narrative, character behaviour and mood lighting. The developed emotion recognition system achieves an accuracy of 98% trained on two million

images, moreover a continuation desire system is developed which achieves 95.1% accuracy trained on 2.6 million samples, when used in a co-op two player game session and 78.5% in a single player game session.

Furthermore, an expert interview with a game company reveals that a tool to measure emotions and continuation desire such as the system suggested in this paper is applicable in the production phase of a game, and can give great value to the developers in early stages as well as later in the production phase.

2 Related Works

Creating adaptive video games is drawing more attention from both players and game studios, as the research have found that stimulating the mind, with signals that create an emotional response is an import component of game design [13,37].

Machine learning have also been used to predict emotions from different kind of physiological signals, such as heart rate, video data, galvanic skin response, electroencephalogram (EEG), eye tracking etc. [59,62]. It is a promising way for seamless and accurate measures of emotions during game play in real time.

When working with different measures, both physiological and video feeds, one has to choose a subjective measure which best fit the way of evaluating the experience of play. Different models exist to evaluate the experience of play, but one model for evaluating engagement as continuation desire, proves as a useful tool for game design and analysis of games [45,48,49]. As of now, the connection between continuation desire and different physiological signals have not been explored in any academic sense to its fullest, apart from [51]. Which used Galvanic skin response, eye tracking and heart rate to determine the level of continuation desire a user was feeling, during game play, the project achieved an respectable accuracy of 75%.

Outside of academia, there are multiple products related to emotion recognition, especially targeted at affective gaming, and product analysis. A common denominator is using physiological signals and a processing method, either in terms of signal processing or using neural networks to model the data to an emotion. Typically in this space, arousal and valence is used to describe each emotion, instead of the actual emotion itself. This is due to the fact that valence and arousal are continuous values which in theory can map any combination of values to a direct emotion, as seen on the circumplex model of affect [41].

Most companies accept different input modalities from heart rate to electroencephalography. The most complete and complex setup is from Sensum [3], where it is possible to provide tailored emotion recognition setups for a specific project, and use almost any kind of input, and output various data. Only one of the companies focus on testing within games and experiences, which is Mod.ai [2], and their products is within the field of player experience and retention. However, they use player behaviour telemetry inside the game, and do not measure any physiological signals. Therefore it seems that there is no company which specifically focuses on player and user experiences in the games or film industry,

by utilising the concept of continuation desire. Based on the current research specific areas are extracted and elaborated upon in more detail in the following literature review, explicitly Engagement, Continuation Desire, Emotions and Machine Learning.

3 Literature Review

3.1 Engagement

Engagement is an element of the player experience which is highly important, and it is of utter importance for games to be engaging to be considered a good game. It is not enough to only motivate players to keep playing, they have to be engaged to keep playing a game [46]. Player engagement is one facet of the player experience, and can be related to flow [10,12], presence [31], immersion [29], enjoyment [27], affective dimensions, and satisfaction [5]. With these dimensions one can fully define and attempt to measure engagement. This can play a large role in evaluating video games, as engagement, enjoyment and immersion arise from a volition to experience the game.

Engagement is thus an intertwined concept and can further be defined as:

“A value of user-experience that is dependent on numerous dimensions, comprising aesthetic appeal, novelty, usability of the system, the ability of the user to attend to and become involved in the experience and the user’s overall evaluation of the salience of the experience.” [35]

In relation to player engagement, the concept of continuation desire is essential to include, as for a player experience to be engaging, the player must have the desire to continue. According to Brown et al. [9], in the context of play, the desire to keep playing is a product of play and that the pleasure of the experience makes a player continue playing. In [46] we set the desire to continue in the context with player engagement, and we elaborate which characteristics of a players engagement that makes a player want to continue playing.

3.2 Continuation Desire

When evaluating engagement as the desire to continue, the term Conation can be used as a foundation to give an overview of the topic. Conation was first defined in the eighteenth century, as one of three parts of the mind: Affection, Cognition and Conation (Continuation Desire) [23]. Conation was then and still is defined as the desire and will to strive for a goal, and as this connection between knowledge and affection which leads us to act [26]. Huit describes conation as the following:

“The personal, intentional, planful, deliberate, goal-oriented, or striving component of motivation, the proactive (as opposed to reactive or habitual) aspect of behaviour” ([26], p.1).

This means that conation is the intrinsic motivation that is displayed when attempting to achieve a goal through volition. In [45,50], we redefine conation in the context of digital media, more explicitly video games as continuation desire, and formalise it as the player engagement process framework (PEP) [46].

Continuation desire is a way to describe a player’s engagement based on different triggers which cause players to engage in or disengage from an interactive experience. The player engagement process presents a comprehensive connection between four components which forms continuation desire: objectives, activities accomplishments and affect [47]. The relationship between these components can be seen on Fig. 1.



Fig. 1. Causes of continuation desire add to the level of conation. Source: [48]

The objective(s) of an interactive experience is the extrinsic and intrinsic motivations to reach a goal or overcome a challenge a player experiences during play. These includes the extrinsic objectives which are set up by the game as well as any intrinsic objectives that a player brings to or forms during the experience.

The activities describe the ways a player can become engaged with the experience while pursuing the objective(s). The accomplishments are defined as receiving achievements, experiencing progression in the narrative or levelling up as well as completing objectives.

The activities, objectives and accomplishments of a game thus comes from either the conscious design of the developer, which have designed the experience, from specific items, narratives, or it can come from the preconceived expectations of the game that players can have, or objectives set up by the players themselves.

Lastly the affect is defined as the emotions which are experienced during play as well as the absorption of the player into some activity. The affect can also be described as the conscious or sub conscious emotional response experienced by the player. This response can cause a physiological, cognitive and behavioural reaction, such as facial expressions, increase/decrease in heart rate and higher/lower conductance with electrodermal activity [8,47].

The affect can then be translated into the positive or negative emotions which makes or breaks the experience, or in other words pull the player in or out of the experience.

These four components of the Player Engagement Process in combination can add to the level of conation a player is currently experiencing.

The components interacts sequentially with each other in cyclical rotation. This cycle is the player engagement process, and the process may begin with the affect of the player. The reason is that the player will initially have a form of intrinsic motivation to start the game before the experience begins.

The experience provides the player with different objectives, which can be accomplished through defined activities. The accomplishments, (or merely performing the activities) can then result in the player becoming absorbed and/or lead to a negative or positive affect which makes the player want to continue or disengage from the experience [46].

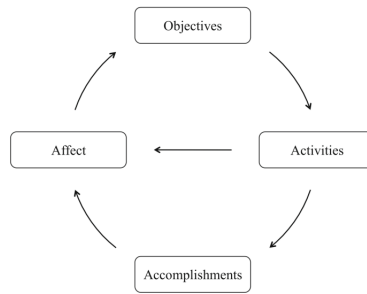


Fig. 2. The relation between objectives, accomplishments, activities and affect (The OA3 framework). Source: [46]

The emotions which relate to affect can cover the whole spectrum of emotional responses, from low to high arousal and low to high valence. These changes in human behaviour can be measured and there is therefore potential to classify them using machine learning [28]. When using such a method it is thus possible to obtain continuation desire without using subjective measures, as the affective space of elements of the continuation desire can be measured through emotions.

This is also supported by [16], who found that there is a significant correlation between physiological measures and self reported measures revolving a players experience in a FPS (First Person Shooter) game.

Continuation desire is a combination of many facets and it can be argued that all actions will lead to a affective response, as seen on Fig. 2, therefore we suggest that it is possible to use the affective state of a player to measure continuation desire.

In summary, the feeling of the desire to continue, can be described by affect or emotions, as a person often react to content with an emotional response. The next section will therefore focus on emotions, and how to measure a persons emotional state and which cognitive processes and state can influence a persons emotional response.

3.3 Emotions

Almost 150 years ago in 1872, Charles Darwin wrote in his book called ‘The Expression of the Emotions in Man and Animals’,

“Facial expressions of emotion are universal, not learned differently in each culture” [14].

Today, this is still a statement that has been much debated. However, psychologist Paul Ekman, proposed six universal emotions, sadness, disgust, fear, anger, surprise and happiness [17, 18, 39].

In both the field of psychiatric and neuroscience research, there is a general consensus that through evolution, humans have been supplied with a range of basic emotions [18, 39]. Each of the emotions is unique in the physiological and behavioural expression, and each of them emerges from an activation within particular neural pathways in the central nervous system.

The emotional system is complex, and it can be described by the emotional continuum, see Fig. 3. A human has three orders of emotion, the first order is the automatic processes within the human body, such as appetites. As it is bodily responses, which is an automatic process, that is hard to control by cognitive processes, and is a responses or reaction to a stimuli and is therefore categorised as uncontrollable responses.

The second order is basic emotions, which as mentioned, are the universal emotions: emotions, sadness, disgust, fear, anger, surprise and happiness. These emotions can also be translated to high/low arousal and positive/negative valence [39].

The third order is higher order emotions which requires cognitive process to be activated, these are emotional responses which are activated by complex neural signals. Higher order emotions can be pride, anxiety, remorse, etc. As these emotional responses are produced by complex signal and cognitive processes makes them significantly more subjective than the basic emotions. The third order is also tertiary emotions, which means that they require higher level cognitive processing, as the emotions are usually self-conscious, and require self reflection and evaluation.

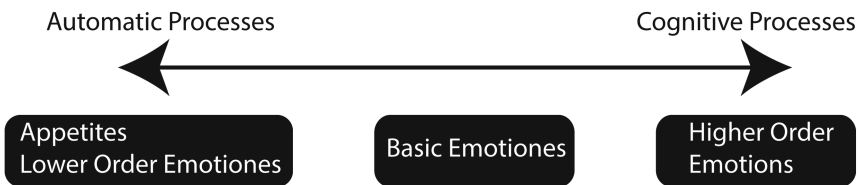


Fig. 3. Figure showing the emotional continuum and the different emotion orders.

There exists six different states, which cover the phenomena concerning the emotional responses, as many emotions and responses is depended on the state

of mind a person [41,42]. The states are: emotion, feelings, moods, attitudes, affective style and temperament. These states could and in many cases will affect the measuring of emotional responses, both with self reported measures or measured with physiological measures [41,42].

As mentioned emotions can be segmented into arousal and valence, and by using the two axes a model can be made to plot emotions on a two dimensional system. Such as model was described by Russell [41], and depicted as the “circumplex model of affect”. The model describes how emotions have both arousal and valence properties which relate to the neural circuitry.

Being able to concise emotions in a visual way and compare emotions on a mapped figure makes it more accessible to compare emotions. According to [39,41,42], is it possible to sub serve all emotions on the model, as valence and arousal cover all affective states.

There has been much research in the difficulties people have in assessing the describing their emotions [43]. When conducting self reported measures of emotions, people tend to either not experience the emotion or recognise emotions when isolated from each other. Cacioppo describes that the psycho physiology as a field that is based on a assumption that emotion and cognition are embodied phenomenon rooted deep in the physical substrate of the body and brain [11]. It must then be possible to measure characteristics of the human body to elucidate and infer the understanding of the structures of third order cognitive functions.

From this section it is found that emotion have roots in the physiological state of mind, and that basic emotions is close to uncontrollable, meaning that it is possible to measure an emotional response. The basic emotions is also referred as universal emotions, which makes it possible for one system to measure emotions from every human on earth, as emotional responses should be largely identical.

Closely related to emotions is the facial expression, and the next section will focus on the visual emotional responses that occurs when a person reacts to stimuli.

3.4 Facial Expressions

As Darwin wrote, all facial expressions are universal, and even the debate that some could be culturally different, the basic emotions is universal to a high degree. From newborns to blind persons to the average person, all show the basic emotions in the exact same way [32]. This study concur with Darwin in some manner, that most facial expressions is universal, especially macro expressions, but micro expressions might have a slight variance depending on the cultural background of a subject. The two different kinds of expressions, macro and micro-expressions, denotes both the duration and the area of the facial musculature that is used. Macro-expressions often last 0.5–4 s and involves the entire face [19]. Macro-expressions is mostly single emotional responses and last said duration when the emotion is not being concealed, by higher cognitive functions. Macro-expressions occur mostly when we are alone or with friends or family, when we feel most open and familiar.

Micro-expressions are however different, as they last as little as 1/30 of a second and involves only small parts of the face. Micro-expressions are therefore arduous to see, especially for a human. They are often categorised as concealed emotions as they occur when two parts of the brain, the cortical motor strip and subcortical areas, oppose over the neural pathways to take control over the facial expression. This ends with quick fleeting micro expressions.

An evaluation test developed by Matsumoto & Hwang concludes that humans can on average classify micro-expressions and subtle macro-expressions with a 48% accuracy, and when both joy and surprise is removed, which are the easiest to recognise, the recognition falls to 35%. This means that most people are not very good at recognising facial expressions, when the emotional response is short.

Several researchers have investigated how high cognitive load affects the response of an emotional expression, and usually high cognitive load tends to suppress emotional responses yielding a more neutral expression or of lower intensity. This means that, in relation to games, that during a challenging level, players might not express high intensity emotions in the moment, or even any. Expressions can be enforced by being in a social setting, when people playing a game together, expressions occur more often than when alone. This means that being in a social setting, amplifies your emotional expressions as the context is more appropriate to express emotional responses [20]. Measuring emotions from facial expressions can introduce challenges as expressions is depended on the environment and setting around the person. It is therefore important to create an environment and setting that facilitate responses in a safe manner.

4 Machine Learning

Classifying and predicting human emotions is a hard task, even for humans. Creating a machine learning algorithm which can predict and classify such emotions is equally hard. As the era of machine learning is upon us, research in this area is moving at enormous speeds, also within predicting and mimicking human behaviour.

When working with images, or image sequences (video), convolutional layers is commonly used to extract features of each image. As images are 2D, having a width and height, 2D convolutional layers is used, which takes a kernel and literally processes the image. The essence of a convolutional layer is to extract feature patterns from a picture. In terms of Facial Expression Recognition (FER), the kernel will learn to look for specific lines in the face, such as eye placement, eyebrow angle, mouth width and angle, as these landmarks usually relates to a specific expression.

A study by T. Zhang et al. used EEG data and a LSTM model to predict emotions with a 90% accuracy [62]. The network was constructed with a quad-directional spatial LSTM layer, after this layer a bi-directional temporal RNN layer is used to spot more subtle emotion patterns, lastly a softmax layer creates probability for each class.

A study using only a video feed for facial recognition to classify emotions, have achieved high accuracy (+80%) [36]. It is stated that it would be possible to

use in real time, but no such tests is conducted in the studies. A common factor is that the dataset is often small, and consists of only a few people, meaning that the algorithm is learning specific responses relating to the person. This means that the algorithm is not learning a generalised way of predicting emotions, which will make the algorithm useless in a real environment. Therefore it is important to have as many different samples as possible in the data set.

Our study bases its algorithm on a shallow but deep architecture, using both residual connections [22, 54], high-way layers [24] and shared parameter states to create a novel architecture which is highly optimised and achieves state of the art results on emotion classification. The developed emotion recognition system was trained on 2 million samples and achieved a high accuracy of 98% on the test set.

The continuation desire algorithm is a sequence based model of LSTMs, as conation is a feeling that is time based. The algorithm will ingest data from the emotion system both in terms of facial embeddings, but also softmax scores. With this rich feature set, it is possible to accurately predict a users conation level. The algorithm was trained on 2.6 million samples and achieved a high accuracy of 95.2% on the test set, which is more than previous work in this area [51].

To get an indication of the speed of the algorithms, the inference time for the emotion recognition system, a single frame takes 0.002s to process making it indeed quick enough to run along side multiple other high load systems, such as a game. For the continuation desire predictor, it processes 18000 samples in 0.1s, as this algorithm runs every five minutes, the slightly larger overhead is not significant, and would still be able to run alongside a game. This means that both algorithm are able to run in real time, even simultaneously with a complex graphical 3D game.

5 Experiment

When designing engaging experiences, levels of conation is a good measure of player experience as the PEP is a cohesive model, rooted in affect. Affect correlates closely with valence and arousal which makes up emotions described by the circumplex model. Therefore if measuring emotions, either with physiological data, or a video feed, it is possible to determine levels of continuation desire through the use of said data. To measure emotions truly unobtrusive, using a camera to capture facial expressions is the optimal way, since other physiological signals such as heart rate and galvanic skin response require equipment attached to the test participant. The setup can quickly become complex and expensive, and the scope of this project is concerned with creating and testing a software system which is accessible and applicable for e.g. smaller development studios.

In terms of facial expressions during high cognitive loads, where faces usually become very neutral, this study is founded on the co-operational play style in two player games to enforce interaction, even through challenging play levels. A neutral face may occur, and this could also relate to conation. The expressions are determined usable to infer conation from said data.

Using faces as features, it is optimal to use machine learning, especially deep learning to extract deep facial features during experiencing digital content, these deep features can both be used for emotion recognition but also to determine the level of conation. This could be a powerful tool for both affective and adaptive gaming, for changing narrative behaviour, but also for game or film evaluation, to investigate the underlying emotions and continuation desire during a game or film. The use such a system for adaptive system, the developed recognition system should be able to process data in real-time, therefore the developed system needs to be lightweight in terms of compute power needed, and usable cross platforms for PC, Unix based systems and mobile operating systems.

Two machine learning algorithms is developed in this study, namely the emotion recognition system, and the continuation desire predictor. The emotion recognition system was developed by using Deep Learning trained on two million images across eight different emotions. The continuation desire predictor system was developed by creating a downstream task on the emotion recognition system, using rich feature vectors. The data for training this system was gathered by pilot testing the game “Little Big Planet”¹ for Play Station 4. The game is a platformer in semi 3D, which can be played in a co-op setting. The test persons played for at least 30 min, and reported their desire to continue every five minutes during play on a scale from 1 to 7. This resulted in 2.6 million data points.

The systems has then been validated to access its accuracy in other domains from where it was developed. This validation was done by using the game “This is the Only Level”², which is a lightweight 2D graphics game where one player have to complete stages of the same level where small mechanics change from stage to stage.

The method for the test will be for test players to complete 15 stages of “This is the Only Level”, after each stage completed the test persona will report their current desire to continue on a scale from one to seven.

Cross validation between the reported continuation desire and the predicted values will be done, to analyse the true accuracy of the system. To analyse the true accuracy of the emotion recognition system, sequences of images of the session will be validated to see if they correspond to the recognised emotion of the system.

The validation test was conducted over three days, all in all 18 participants were tested ($n = 18$), resulting in a mean play length of 8.2 min and a variance of 1.43 $x = 8.2, \sigma = 1.43$. This aggregates to 147 min of game play video. All participants completed all 15 stages of the game.

The age distribution of all participant had an age range from 22–30 $x = 25.61, \sigma = 1.94$. The sampling method for the test is convenience sampling. This means that the distribution of ages is much the same as in the continuation desire test session. Most participants were about 25–26 years of age, while a couple of participants were in their early twenties or late twenties.

¹ <https://www.playstation.com/en-us/games/littlebigplanet-3-ps4/>.

² John Cooney, 2009 (jmtb02) <http://www.jmtb02.com/this-is-the-only-level>.

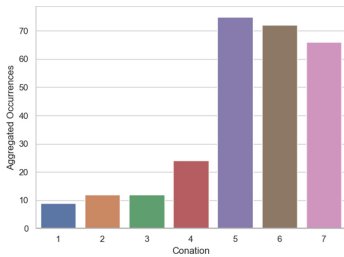


Fig. 4. Continuation desire levels reported from the test on the game

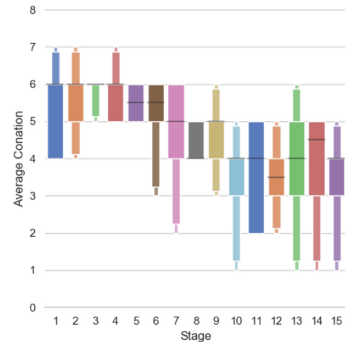


Fig. 5. Reported continuation desire at each stage from the game test.

The participants also reported the weekly hours spent on gaming. The reported average was 9 h per week and this indicates that the general skill level of the participants is higher than the average, so it was expected that most participants would complete all stages in a quickly manner.

On Fig. 4, the aggregated occurrences of the reported continuation desire levels can be seen (x axis/Conation levels: 1 = low, 7 = high). The aggregated occurrences of conation/continuation levels in the test of “this is the Only Level” ranges from 1 = low, 7 = high. In the test, the conation levels created a negative skewed distribution being that over 70% of the occurrences were above level 4.

On Fig. 5 a box plot show the mean continuation desire values reported from each participant. It shows all 15 stages at the x-axis, and there is a tendency for the continuation desire to decrease over stages. This is assumed as the game is quite repetitive, and can become boring after several levels. It is also seen that in the first seven stages the reported continuation desire is stable, and does not contain much variance. Towards to end stages, the variance becomes larger, and here the players preferred game genre might come in to play. Some players do like the repetitive game play, where others do not, and if you are not too fond of the repetitive game style, the game will become boring, and thus feel disengaging. Overall it can be seen that the level of continuation desire decreases over time.

During the test, three emotions was not spotted by the algorithm, namely surprise, calm and disgust. It is not surprising that these emotions were not captured, as the game play does not actively stimulate those emotions. Happy and sad were the most captured emotions by 80%. Neutral and angry accounted for 10% each, while fearful was 2%.

The sad response is surprising as the game does not create situations where a sad response makes sense. The distribution however is different from the emotion distribution from the pilot test on “Little Big Planet”, where the most captured emotion was angry (as players were sometimes frustrated).

The emotion recognition system was validated by taking every 30th frame from the video taken of each participant and manually tagging the frame with the emotion which the validator believes is fitting. The whole test have 132,387 frames with faces, every 30th of that is 4,412 frames which have been analysed, to see if the recognition algorithm predicted correctly. The investigation revealed that the emotion recognition system predicted correctly 91.28% of the time, making the algorithm well generalised from the training. It was also noticed that, as mentioned in the 1st iteration result section, that the background, light and shadows was producing significant noise in the predicting, indicating that the algorithm is sensitive to the environment in which it is operating.

Moreover as sad is also identified almost as much as happy, it was noticed that the facial expressions when sad was predicted, looked more like it was supposed to be neutral or frustration over the current stage in the game.

On Fig. 6, the graph shows the play-through analysis of the 15 stages of “This Is The Only Level” for one participant. In this particular session, the continuation desire predictor have mostly low continuation desire through out the session, but the participant had high continuation desire the first seven stages (until around frame 1000) and thereon after low continuation desire. It also shows two strong periods of happiness, in between the two spikes, sadness rises. In this sadness spike, the participant had problems with completing a level. The participant looked frustrated which at times can look like a sad or angry facial expression, this frustration is something that the emotion recognition algorithm does not account for, and therefore subjects can be classified wrongly in certain situations.

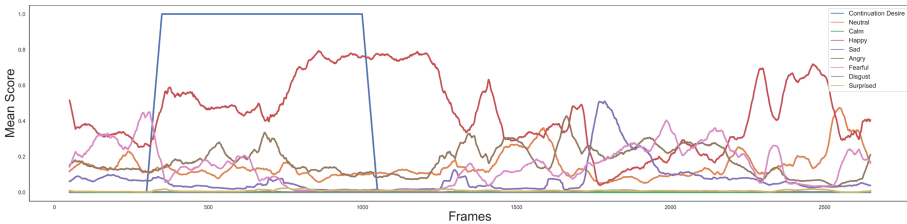


Fig. 6. Graph showing a play through analysis for all 15 stages from the validation test on the game “This is the only level”, the y-axis show the aggregated probability of the emotions. The x-axis depicts frames after start (1 frame = 1/30 s).

To analyse how well the continuation desire predictor performed on the test, the reported values were compared to the predicted values. A comparison of the percentages of low/high continuation desire level from the algorithm and the participants’ own reported level was conducted. The percentages from the algorithm’s levels are depicted in Fig. 7 and for the reported levels from the participants in Fig. 8. The figures were based on the game “This is the only level” and made by calculating the percentages on high/low continuation desire

of both the reported and predicted levels. The reported levels from participants were converted from seven levels to high/low by binning them. Continuation desire level 1, 2, 3, 4 is low, while 5, 6, 7 is high. In the comparison analysis the algorithm predicted correctly 78.48% of the time, which is lower than observed during training on the test set. One reason why the accuracy is lower is that the domain in which the algorithm was used is not where it was meant to be used, as this is not a co-operational game or multiplayer game as in the pilot test. This means that there is no interaction between players which does increase a subject's will to express emotional responses. Taken that into account the algorithm does perform quite well for a real scenario even if the applied environment is out of context.

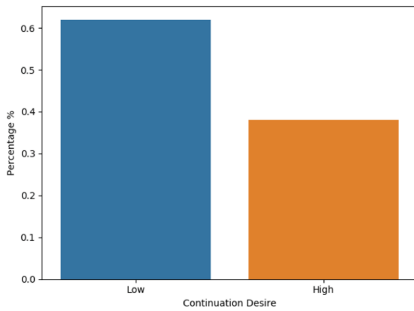


Fig. 7. Barplot showing the algorithm's predicted continuation desire percentages from the test.

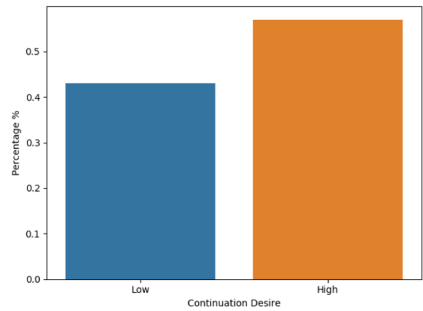


Fig. 8. Barplot showing the reported high/low continuation desire percentages by participants from the test.

It can be seen that the algorithm does predict less 'high' level of continuation desire relative to the true reported level from the subjects (38% high predicted versus 58% high reported), this can be explained with that the subjects had more periods with neutral or sad, which the continuation desire predictor takes as low continuation desire. This supports the hypothesis about that the subjects does not express as many reactions when playing single player games. This is also discussed in the literature review, where it is mentioned that when a person is alone, emotional expression is far more uncommon than when interacting with other persons. Moreover the game is also a puzzle game, which can introduce high cognitive load, which does suppress facial expressions as well.

6 Discussion

In the following section a discussion of the study as a whole, what results and new knowledge have been created, and what could have been done to enhance the method, development and test. The goal of the study is to investigate to

which extend it is possible to infer a players continuation desire and emotions during an interactive experience. This had to be done by only using a video feed, and do this analysis in real time as well.

The developed emotion recognition system achieved a high accuracy of 98% on validation test set which is, compared to other algorithms in the research [6, 21, 34, 40, 53, 55], a significant increase to a point where it is (at least theoretically) significantly better than humans. Humans can recognise micro-expressions 47% of the time [1, 32], and macro-expression to a degree of 84% [1, 32] in real life scenarios. The algorithm was validated in the 2nd iteration, and ended at a 91.28% accuracy in a real world test, which is excellent performance compared to 80.17% [34], 86.9% for [53], 78% for [55], 66% for [6], 71.16% for [21], 88.8% [4] and 74.15% for [40]. For the mentioned papers, they all have data quality and sufficient data set in common, as mentioned in the analysis many papers use data set which consist of less than 1000 samples, and with even fewer different persons, this means that the algorithms learns the data, and only the persons in the data set. This is a significant bias, which is common for papers that are not published recently. This introduced a bias to the accuracies, meaning that the proposed earlier solutions only worked on internal data sets, and not in real scenarios. The current study has used 2,000,000 images on 500,000+ different persons, which makes the final results more valid than tests on smaller data sets. However, the results between studies are not directly comparable as no benchmark data set has been made for emotion recognition as of yet, and this is a problem for comparing different approaches and algorithms. This is a common practise in other fields within machine learning, such as natural language and image processing.

During development of the emotion recognition system, it was observed that image quality and lighting could interfere with the algorithms predictions. To account for this, more work needs to be done with the pre-processing of the input features, so light and shadows do not bias the image to such a degree that facial expressions are being mis-classified. A thorough investigation should have been conducted on data from the data gathering test to see the actually cause of the errors.

The data used for the emotion recognition system, is of high quality, but observed from the data, many of the expression are of high intensity, meaning that some expression are over exaggerated to a degree that would not occur in a real environment, this was mostly from the FER data set, which had 35,000 samples [21]. A data set formed from playing subjects emotional responses would be of greater value and possibly create a more specialised algorithm for interactive experience emotion recognition.

The continuation desire algorithm achieved high accuracy as well, 95.2% was the final accuracy on the two player pilot test set from training, which is higher than previous work done in this area. In the out of context testing of the single player game "This is the Only Level", an accuracy of 78.48% was achieved. Because the algorithm was trained on data from the data gathering test from the 1st (two player) iteration, where bias in the test environment have made

the data biased in terms of predicted emotions, the algorithm may have learned wrong underlying relations between emotions and continuation desire. This error in the data could implicate the continuation desire algorithm to an extent where the output is inaccurate. But the real world test on “This is the Only Level” shows that it still performs relatively well in an out-of-sample environment to an extent where it is acceptable, but it is not as close to the original pilot test set accuracy as seen with the emotion recognition algorithm. Further testing needs to be done to verify the in-sample environment accuracy of the continuation desire algorithm, as it is predicted that it will perform better on a multiplayer or interactive game.

The continuation desire algorithm is trained on only 12 subjects’ data, which is to a degree sufficient, but with more data, the algorithm would be able to accommodate more player types and expression styles. More data would also make the algorithm more robust to different emotional changes which happens in games, as games have different flows and tempo which affects the emotional response. Capturing this data is time consuming but it is needed to create an even better performing algorithm. Using one type of game also comes with risk, as each game produces different emotional curves, testing on different games should also be done to minimise the classification bias which is introduced by only using one game.

The eight emotions that have been used as output from the emotion recognition model, was chosen because of the basic emotions, happiness, fear, sad, surprise, anger and disgust, these emotions are the most occurring ones, but humans are not expressing emotions at all times, therefore neutral and calm was added to accommodate this behaviour. However, from investigating hours of video data from game playing, it becomes more apparent that those expressions might not be enough, as frustration and focus are facial expressions, that are very frequent during game play, especially in single player games, and this typically means that the player is engaged. But these emotions are very close to other emotions as well, such as frustration to sadness and anger, and focus to calm and neutral, distinguishing between these expressions are difficult. This could possibly be done by another approach such as outputting arousal/valence values instead, so that each emotion has a numerical value, which could create more detailed incremental steps between emotions. An intensity index on each emotion could also enhance the interpretability of the predicted emotion, to see if the person is very angry or less angry.

Gathering these two algorithms and using them to infer emotions and continuation desire from video data was done in a case study with the indie game company Invisible Walls based in Copenhagen. The overall points from the interview were that although the company was still in early stages of their development of their next game, they could see that a data driven approach like the one suggested in this study could give great value in their tests, even in their short usability tests. Due to the only requirement for the system being a video camera, no additional equipment for an expensive and complex setup is needed, it becomes attractive to smaller studios as well. This use case is only based

on one interview with one game studio, but the information gathered leads to larger game studios benefiting from a system like this as well. This approach is also largely seen in the marketing sector where emotion recognition is used extensively to see responses from test groups regarding a commercial or product.

In summation, the developed algorithms and evaluation framework proves that it is indeed possible to predict emotions and continuation desire in real time using only video. The system is usable in its core, and with different tweaks and additional features it could work as a method for testing and evaluation of a variety of interactive experiences.

7 Conclusion

The purpose of this study was to investigate the possibility to measure and classify emotions and continuation desire during game play from a video stream in real time by using state of the art and novel machine learning approaches. The proposed solution is an emotion recognition system which perform better than state of the art algorithms and (theoretically) humans which achieves 48% on facial expression recognition tasks [33]. Furthermore the system is trained on the currently (as of June 2019) largest publicly known emotion recognition data set.

The results of this study is an emotion recognition algorithm which performs in static test sets 98% accuracy and in a real two-player scenario 91.28%, which is quite high, even with the problems in the emotion recognition data. A continuation desire algorithm which performed 95.2% of accuracy and 78.48% in an out of environment test with a single player game, suggesting an algorithm which is well generalised and use-able on different game genres and styles. The case study furthermore shows that such a system is also usable in a game company.

It can therefore be concluded that we successfully created an emotion recognition and continuation desire predictor algorithm which is usable in real time and usable for game testing which creates value for game testing teams during development. With minor changes the algorithms could perform better and achieve higher accuracies, while making them more robust to changes in the environments surrounding the participants.

References

1. Nonverbal Communication: Science and Applications AU - Matsumoto, David AU - Hwang, Hyi Sung. SAGE Publications, Inc., Thousand Oaks (2013). <https://doi.org/10.4135/9781452244037>. <http://sk.sagepub.com/books/nonverbal-communication>
2. Modl.ai (2019). <http://modl.ai/>
3. Sensus (2019). <https://sensus.co/>
4. Albanie, S., Nagrani, A., Vedaldi, A., Zisserman, A.: Emotion recognition in speech using cross-modal transfer in the wild. arXiv preprint [arXiv:1808.05561](https://arxiv.org/abs/1808.05561) (2018)

5. Amir Zaib Abbasi, D.H.T., Hlavacs, H.: Engagement in games: developing an instrument to measure consumer videogame engagement and its validation. *Int. J. Comput. Games Technol.* **2017**, 1–10 (2017)
6. Arriaga, O., Valdenegro-Toro, M., Ploger, P.: Real-time convolutional neural networks for emotion and gender classification. *CoRR abs/1710.07557* (2017)
7. Bitouk, D., Verma, R., Nenkova, A.: Class-level spectral features for emotion recognition. *Speech Commun.* **52**(7), 613–625 (2010)
8. Brett, A., Smith, M., Price, E., Huitt, W.: Overview of the affective domain. *Educ. Psychol. Interact.*, 1–21 (2003)
9. Brown, S., Vaughan, C.: *Play: How it Shapes the Brain, Opens the Imagination, and Invigorates the Soul*. Avery (2009). <https://books.google.dk/books?id=ESQDsgqfgusC>
10. Buchanan, R., Csikszentmihalyi, M.: Flow: the psychology of optimal experience. *Des. Issues* **8**(1), 80-1 (1991)
11. Cacioppo, J.T., Tassinary, L.G., Berntson, G.: *Handbook of Psychophysiology*. Cambridge University Press, Cambridge (2007)
12. Chen, J.: Flow in games (and everything else). *Commun. ACM* **50**, 31–34 (2007)
13. Christy, T., Kuncheva, L.I.: Technological advancements in affective gaming: a historical survey. *GSTF Int. J. Comput. (JoC)* **3**(4), 32–41 (2013)
14. Darwin, C.: *The Expression of the Emotions in Man and Animals*. D. Appleton and Co., New York (1872)
15. Debeauvais, T.: *Challenge and Retention in Games*. UC Irvine. ProQuest ID: Debeauvais_uci_0030D_13948. Merritt ID: ark:/13030/m53n6r1p (2016). <https://escholarship.org/uc/item/6k3357qx>
16. Drachen, A., Nacke, L.E., Yannakakis, G., Pedersen, A.L.: Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In: *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, pp. 49–54. ACM (2010)
17. Ekman, P.: Body position, facial expression, and verbal behavior during interviews. **68**, 295 (1964). <https://doi.org/10.1037/h0040225>
18. Ekman, P.: An argument for basic emotions. *Cogn. Emot.* **6**(3–4), 169–200 (1992)
19. Ekman, P.: Emotions revealed. **12** (2004). <https://doi.org/10.1136/sbmj.0405184>
20. Frith, C.: Role of facial expressions in social interactions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**(1535), 3453–3458 (2009)
21. Goodfellow, I.J., et al.: Challenges in representation learning: a report on three machine learning contests. In: Lee, M., Hirose, A., Hou, Z.-G., Kil, R.M. (eds.) *ICONIP 2013. LNCS*, vol. 8228, pp. 117–124. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-42051-1_16
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. *CoRR abs/1512.03385* (2015). <http://arxiv.org/abs/1512.03385>
23. Hilgard, E.R.: The trilogy of mind: cognition, affection, and conation. *J. Hist. Behav. Sci.* **16**(2), 107–117 (1980)
24. Huang, G., Liu, Z., Weinberger, K.Q.: Densely connected convolutional networks. *CoRR abs/1608.06993* (2016). <http://arxiv.org/abs/1608.06993>
25. Hudlicka, E., Broekens, J.: Foundations for modelling emotions in game characters: modelling emotion effects on cognition, pp. 1–6 (2009). <https://doi.org/10.1109/ACII.2009.5349473>
26. Huitt, W., Cain, S.: An overview of the conative domain. *Educ. Psychol. Interact.*, 1–20 (2005)
27. Ijsselstein, W., et al.: Measuring the experience of digital game enjoyment. In: *Proceedings of Measuring Behavior* (2008)

28. Jang, E.H., Park, B.J., Kim, S.H., Eum, Y., Sohn, J.H.: Identification of the optimal emotion recognition algorithm using physiological signals, pp. 1–6, November 2011
29. Jennett, C., et al.: Measuring and defining the experience of immersion in games. *Int. J. Hum.-Comput. Stud.* **66**(9), 641–661 (2008)
30. Lee, C., Yoo, S., Park, Y., Kim, N., Jeong, K., Lee, B.: Using neural network to recognize human emotions from heart rate variability and skin resistance. In: *Engineering in Medicine and Biology 27th Annual Conference* (2005)
31. Lombard, M., Ditton, T.: At the heart of it all: the concept of presence. *J. Comput. Mediat. Commun.* **3**(2), 0 (1997)
32. Matsumoto, D., Willingham, B.: Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *J. Pers. Soc. Psychol.* **96**, 1–10 (2009). <https://doi.org/10.1037/a0014037>
33. Matsumoto, D., Hwang, H.S.: Reading facial expressions of emotion. *Psychol. Sci.* **115**, 541–558 (2011)
34. Moore, S., Bowden, R.: Local binary patterns for multi-view facial expression recognition. *Comput. Vis. Image Underst.* **115**(4), 541–558 (2011)
35. O'Brien, H.L., Toms, E.G.: What is user engagement? A conceptual framework for defining user engagement with technology. *J. Am. Soc. Inform. Sci. Technol.* **59**(6), 938–955 (2008). <https://doi.org/10.1002/asi.20801>
36. Ouellet, S.: Real-time emotion recognition for gaming using deep convolutional network features. *CoRR abs/1408.3750* (2014)
37. Picard, R.W.: Emotion research by the people, for the people. *Emot. Rev.* **2**(3), 250–254 (2010)
38. Pollreisz, D., TaheriNejad, N.: A simple algorithm for emotion recognition, using physiological signals of a smart watch. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2353–2356 (2017)
39. Posner, J., Russell, J.A., Peterson, B.S.: The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **17**(3), 715–734 (2005)
40. Rashid, M., Abu-Bakar, S.A.R., Mokji, M.: Human emotion recognition from videos using spatio-temporal and audio features. *Vis. Comput.* **29**(12), 1269–1275 (2012). <https://doi.org/10.1007/s00371-012-0768-y>
41. Russell, J.A.: A circumplex model of affect. *Pers. Soc. Psychol.* **39**, 1161 (1980)
42. Russell, J.A., Barrett, L.F.: Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J. Pers. Soc. Psychol.* **76**, 805 (1999)
43. Saarni, C.: *The Development of Emotional Competence*. Guilford Series on Social and Emotional Development. Guilford Publications, New York (1999)
44. Sagi, B.R., Silvestrini, R.: Application of combinatorial tests in video game testing. **29**, 745–759 (2017). <https://doi.org/10.1080/08982112.2017.1300919>
45. Schoenau-Fog, H.: Hooked! – evaluating engagement as continuation desire in interactive narratives. In: Si, M., Thue, D., André, E., Lester, J.C., Tanenbaum, J., Zammitto, V. (eds.) *ICIDS 2011*. LNCS, vol. 7069, pp. 219–230. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25289-1_24
46. Schoenau-Fog, H.: The player engagement process - an exploration of continuation desire in digital games. In: *DiGRA - Proceedings of the 2011 DiGRA International Conference: Think Design Play*. DiGRA/Utrecht School of the Arts, January 2011
47. Schoenau-Fog, H.: At the Core of Player Experience: Continuation Desire in Digital Games, pp. 388–410. Wiley (2014). <https://doi.org/10.1002/9781118796443.ch14>

48. Schoenau-Fog, H.: Designing and evaluating conative game-based learning scenarios. In: Busch, C. (ed.) Proceedings of The 8th European Conference on Games Based Learning – ECGBL 2014, pp. 512–518. Academic Conferences and Publishing International, October 2014
49. Schoenau-Fog, H., Birke, A., Reng, L.: Evaluation of continuation desire as an iterative game development method. In: Proceeding of the 16th International Academic MindTrek Conference, MindTrek 2012, pp. 241–243. ACM, New York (2012). <https://doi.org/10.1145/2393132.2393185>
50. Schoenau-Fog, H., Louchart, S., Lim, T., Soto-Sanfiel, M.T.: Narrative engagement in games—a continuation desire perspective. In: FDG, pp. 384–387 (2013)
51. Selvig, D., Stovring, N., Sjoblom, A., Korsholm, J.: Continuation desire and its physiological underpinnings - unpublished (2018)
52. Seng, K.P., Ang, L., Ooi, C.S.: A combined rule-based and machine learning audio-visual emotion recognition approach. *IEEE Trans. Affect. Comput.* **9**(1), 3–13 (2018)
53. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis. Comput.* **27**(6), 803–816 (2009)
54. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning. CoRR abs/1602.07261 (2016)
55. Tawari, A., Trivedi, M.M.: Face expression recognition by cross modal data association. *IEEE Trans. Multimedia* **15**(7), 1543–1552 (2013)
56. Tawari, A., Trivedi, M.M.: Audio visual cues in driver affect characterization: issues and challenges in developing robust approaches, pp. 2997–3002 (2011). <https://doi.org/10.1109/IJCNN.2011.6033615>
57. Tokuno, S., et al.: Usage of emotion recognition in military health care, pp. 1–5 (2011). <https://doi.org/10.1109/DSR.2011.6026823>
58. Vyvey, T., Castellar, E.N., Van Looy, J.: Loaded with fun? The impact of enjoyment and cognitive load on brand retention in digital games. **18**, 72–82 (2018). <https://doi.org/10.1080/15252019.2018.1446370>
59. Wang, C., Pun, T., Chanel, G.: A comparative survey of methods for remote heartrate detection from frontal face videos. *Front. Bioeng. Biotechnol.* **6**, 33 (2018)
60. Xerfi: Global video game turnover between 2011 and 2019 (2019). <https://www.statista.com/statistics/862278/global-video-game-revenues-worldwide/>. Accessed 19 Mar 2019
61. Yu, S.N., Chen, S.F.: Emotion state identification based on heart rate variability and genetic algorithm, pp. 538–541, August 2015. <https://doi.org/10.1109/EMBC.2015.7318418>
62. Zhang, T., Zheng, W., Cui, Z., Zong, Y., Li, Y.: Spatial-temporal recurrent neural network for emotion recognition. *IEEE Trans. Cybern.* **49**(3), 839–847 (2018)