AN ENGAGEMENT ASSESSMENT FOR GAME DESIGNERS AND GAME RESEARCHERS

by

Ruiming Wang

ABSTRACT OF THESIS

Submitted in partial fulfillment of the requirements

for the degree of Master of Science in Game Science and Design

in the Graduate School of the College of Arts, Media and Design of

Northeastern University

April, 2019
Abstract

Getting feedback from users is the fundamental way for game designers to obtain insight for their game. However, the traditional methods of getting feedback in games have its limitations. Since users are often occupied with the game, it’s hard for game designers to get real-time, objective opinions from the player.

This goal of this research is to build a model that take the GSR signal and postures as input and perform a real-time engagement evaluation that is going to help game designer capture the moment of dissatisfaction.
## Contents

1 Introduction ........................................... 7

2 Background ........................................... 9
   2.1 Traditional Game Research Method .............. 9
       2.1.1 Questionnaire .............................. 9
       2.1.2 Facial Expression Coding .................. 10
       2.1.3 Think Aloud ............................... 10
       2.1.4 Focus Group ............................... 11
   2.2 Biometrics In Game Research .................... 11
       2.2.1 Galvanic skin response .................... 12
       2.2.2 Posture .................................. 13

3 Hypothesis Regards to Engagement .................... 14
   3.1 The Motivation of the Research ................. 14
   3.2 The Hypothesis ................................ 14
   3.3 Metrics Used In The Research .................. 15
       3.3.1 Skin conductance response(SCR) ............ 15
       3.3.2 Eye distance .............................. 15

4 Pilot Study ........................................... 16
   4.1 Pilot Study Methodology .......................... 16
       4.1.1 Equipment and Experiment Environment ...... 16
       4.1.2 Experiment Protocol ....................... 18
   4.2 Pilot Study Result .............................. 19
5 Main Study Methodology

5.1 Rapid Affective Game Response Collection (RAGRC) 21
  5.1.1 Establishing ground truth 21
  5.1.2 Solution 22
  5.1.3 The Games Used In The Experiment 25

5.2 Equipment and Experiment Environment 27

5.3 Eye Tracker 27

5.4 Galvanic Skin Response Sensor 27

5.5 Experiment Protocol 29

6 Main Study Results

6.1 Data Overview 32

6.2 Data Prepossessing 34
  6.2.1 Galvanic Skin Response Sensor Data 34
  6.2.2 Eye distance data 34

6.3 Data Analysis 36
  6.3.1 Direct Comparison Between Two Categories Of Gameplay Experiences 36
  6.3.2 Mann-Whitney U Test 37
  6.3.3 Mixed-Modality 38

7 Discussion

7.1 Findings 40
  7.1.1 Galvanic Skin Response 40
  7.1.2 Eye Tracking data 40

7.2 Obstacle to build a mixed modality model 41
  7.2.1 The Numerous Personality And Limited Data Set 41
7.2.2 GSR ......................................................... 41
7.2.3 Head Peak Detection ................................. 41
7.2.4 Ground Truth Not Accurate Enough .......... 42
7.2.5 Data Process Method And Lack of Feature .... 42
7.3 Long Short-Term Memory Based Recurrent Neural Network .... 42
  7.3.1 Advantage ................................................. 43
  7.3.2 Challenge ............................................... 43

8 Conclusion ................................................... 44
1 Introduction

Playtesting is a fundamental way for game designers to get feedback for their game. There are multiple approaches for game designers to get insight from their game. Common approaches include getting direct feedback from users and data analysis[11].

Getting feedback from the players is the most direct way, however, the feedback can be very unreliable at times[Insert References]. Players are having a hard time giving feedback because they can’t remember all the gameplay process, especially the minor details. There are also playtesting method like think-aloud[17] but is also distracting.

On the other hand, data science is based on a huge amount of data generated after the game is released to the market. This kind of analysis requires a large player base and are not very responsive since it requires a lot of players to play the game first. Which means game designers can’t adjust the game until a lot of people already played the game.

Biometric is the science of extracting insight of a person’s emotion with attributes such as physical attribute, chemical attribute or behavioral attribute[2] of the person. There is a lot of research to evaluate the emotion of a person with biometric data.[8]. Some of the attributes can objectively reflect the person’s emotion and require very less effort and dictation from the person, which makes it a promising option for getting insight from the player.

Although those insights derived from biometrics may not be as direct as a verbal response since the game designers have to look into the data and context to understand the feeling, the biometric is good at finding out the subtle negative feeling that even players themselves can’t find. A study about identify valence and arousal with biometric[14] shows that biometric can detect small yet subtle changes in valence that cannot be observed by people themselves.
The goal of the project is to build a mixed modality system that can automatically perform engagement assessment without biometric signal without manually going through the context.
2 Background

2.1 Traditional Game Research Method

Game user research has become an important area for game industry since it helps game designers improve the player experience to make the game more pleasurable and meet designers’ goal[11]. Some of the most common and effective methods are surveys, facial expression coding, continuous annotation, think aloud, and focus group. These approaches have both benefits and limitations.[11][20]

2.1.1 Questionnaire

The usage of questionnaires is one of the most common methods for game user research. It is cheap, and do not require too much effort to gain both qualitative and quantitative data.

The Game Engagement Questionnaire (GEQ) is one of the most widely used tools to measure the levels of engagement specifically elicited while playing video games[7]. Brockmyer’s study invited 107 male undergraduate students to complete a GEQ with respect to their typical video game-playing experience, and another GEQ after playing a first-person shooter game experience. The result showed that a questionnaire is a promising tool for examining risk and protective factors for negative game impact.

Although questionnaires are a very convenient tool for user research, it lacks the depth of the objectivity measures and does not work well when only has a small number of people are available for testing[23].
2.1.2 Facial Expression Coding

Facial expressive behavior is interest to many researchers. There are a variety of methods to measure facial expression. Ekman and Friesen[24][25] developed the Facial Action Coding System to provide a comprehensive assessment of all visible facial muscle movements.

Sawyer, Smith, Rowe, Azevedo, and Lester[29] used facial expression to enhance the user experience in a game-based learning environment. The students played an educational game CRYSTAL ISLAND, and videos were captured of their facial expressions. They created two models based on the facial expression data and gameplay data to predict students’ learning and presence. The facial expression data were extracted through a biometric tracking system, iMotion. The iMotion system uses Facial Action Coding System to extract facial features, including joy, sadness, surprise, fear, anger, disgust, and contempt. Their research found that facial expression tracking can help predict student learning outcomes and student engagement.

The facial expression coding system may not reliable in all kinds of people. It remains uncertainty while applying this system on ethnically, culturally, and aging diverse sample.[3]

2.1.3 Think Aloud

The think-aloud protocol was developed by Lewis[17]. It requires players to talk about what they are thinking while playing the game. Researchers can combine players’ behavior with their descriptions to draw a more effective conclusion in the analysis.

Tan, Leong, and Shen[30] used physiological data and think-aloud to investigate players’ experiences. The participants invited to the experiment were asked to play a first-person puzzle game for 30 minutes, and then verbalize their prior experience
while watching a replay video of their gameplay. They generated 19 think-codes based on the think-aloud data. They found that think-aloud data can uncover some undetected experiences, but also can contradict to the physiological data.

Think-aloud requires some skill from the player since talking what one is doing is not natural for many people, they should have some practice section before the experiment. This method are tend to be distracting since participants can’t focus on the content they are engaging entirely thus made the stimulus less powerful.

2.1.4 Focus Group

Focus groups are another way to gain qualitative data, where researchers gathered a group of players to talk about their opinions and feedback on the game[27].

Poels, De Kort, and Ijsselsteijn[27] organized four focus groups with different game frequency to explore digital game experience. The participants could freely talk and interact with each other about their game experience, including what occasions they start gaming, what experience while gaming, and what experience after gaming. The participants wrote down their experiences after discussion. They gave a comprehensive categorization of game experience, including enjoyment, flow, imaginative immersion, sensory immersion, suspense, competence, negative affect, control, and social presence, based on the focus group discussions. The limitation of focus groups is that there will be bias since people may pick up other’s opinions while talking in the crowd. A couple of strong voices may dominate a focus group and preventing other people to present their opinions.

2.2 Biometrics In Game Research

Biometrics is the technical term referring to the automatic identification of a person based on his or her distinctive characteristics[3]. It has been widely used in a variety
of applications ranging from border crossings to visiting Disney World\cite{4}\cite{2}. There are many popular biometric modalities, including faces, fingerprints, voices, etc.

### 2.2.1 Galvanic skin response

Galvanic skin response (GSR) measures skin conductance as a proxy for psychological arousal. Galvanic skin responses have a slight delay due to the structure of human bodies, but is less sensitive to noise and less ambiguous than facial muscle and heart activity\cite{15}.

Tijs, Brokken, and IJsselsteijn\cite{31} created an emotional adaption game based on multiple player emotion data including blood volume pulse (BVP), skin conductance (SCL), respiration (RSP), etc. Participants played an adapted version of the PC-game Pacman. The player stayed in the same level but the speed level changed during the experiment. Researchers gathered emotion data, demographics, and gaming experience by physiological equipment and questionnaires. GSR data was extracted using the Biotrace+ detection and additional filtering algorithms. The researchers found correlations between valence reports and the emotion-data features. These biometric modalities can be used to distinguish between a boring, frustrating and enjoyable game experience.

Nacke, Kalyn, Lough, and Mandryk\cite{16} used GSR to enhance game interaction. They recruited ten participants played three versions of a single-player 2D side-scrolling shooter game, each version feature a different control scheme. The first control scheme is using directly controlled sensors (EMG), the second using indirectly controlled sensors (GSR and EKG), and the last version require player using a gampad only. The results showed that the participants thought the physiological controls were enjoyable and worked effectively. Although participants said that indirect physiological control (GSR) has a slow response time, this physiological control...
method could be used to affect slow-changing environmental variables of the game as a dramatic or aesthetic device.

However, skin conductivity response can be triggered by multiple emotions, those emotions can be very different. The occurrence of skin conductivity response can only indicates changes in emotions. As for what the change actually is, researchers must to look into the context analysis the experience by themselves.

2.2.2 Posture

Body language is a essential approach for a human to express their feeling. There is a study trying to identify engagement of students with posture[12], a study using a depth camera to captures posture to learn engagement when gaming[10]. There is a clear correlation between posture and engagement. With the help of computer vision, this study will be using eye trackers to measure engagement.
3 Hypothesis Regards to Engagement

3.1 The Motivation of the Research

There are apparent correlations between galvanic skin response and engagement. However, the challenge game researcher facing is engagement is not the only element that has correlations with those biometrics. For example, a skin conductance response can represent not only a positive feeling but also fear[21], stress[18] and so on.

It is hard for the game researcher to understand what skin conductance response (SCR) could actually mean unless they look into the context. The process of understanding the context can be very time consuming and it makes it harder for conducting quantity research with skin conductance response.

It is even harder for game researchers who want to create a automatic system that asses engagement with skin conductance response since its hard for a computer program to understand the context of the experience when a response occurs.

In order to create a system that can identify the difference between skin conductance responses, we will introduce postures and conduct a multi-modality research. There are already study using multi-modality model on affect[26][22]. So the strategy of looking at how posture and galvanic skin response can be used to create multi-modality system that let two metrics compensate each other for the lack of context can be successful.

3.2 The Hypothesis

The hypothesis is by combining galvanic skin response and gesture, we can create an automatic system that can asses engagement without the context gameplay itself.
3.3 Metrics Used In The Research

3.3.1 Skin conductance response (SCR)

Skin conductance response is being used for a long period of time in user research and psychological testing. There are matured mathematical solution for detecting stimulus based on the raw signal input.

3.3.2 Eye distance

There are researches that indications that postures can be an indication of engagement. The metrics for posture used including depth camera, coding for video recording and so on. A metric that can be easily interpreted by a automatic system is needed and eye distance can be an ideal metric for this subject.

Due the progress made in the of computer vision, eye tracking are becoming more and more affordable and reliable. Also, compares to the position of the head or actual postures, the eye distance is very easy to unify since no matter what postures persons are used to, they have to use their eye to engage with the content. Thus eye distance can be a powerful metric that bypasses personal seating habit.

However, there are few studies that indicate there is an actual correlation between engagement and eye distance, so a pilot study is conducted to prove that the correlation does exist and eye distance can be used as a metric to build the system.
4 Pilot Study

In order to investigate the hypothesis of this thesis, two experiments were conducted: a pilot study and the main study. The pilot research is conducted prior to the main research, and the goal of the research is to find if there is a correlation between the engagement and the posture of players playing a video game and if eye distance can represent the posture effectively. The conclusion drawn by the pilot research inspired the main research.

4.1 Pilot Study Methodology

In this section, we describe the methodology applied in the pilot study.

4.1.1 Equipment and Experiment Environment

The participant is seated in front of a monitor in a soundproofing room. An eye tracker is attached to the bottom of the screen and a cell phone with audio recording is placed on the desk under the monitor.

Eye Tracker  An eye tracker is attached to the bottom of the monitor. (Figure 1)

Eye tracking has been a well-known and vastly used method for studying attention of individuals. The eye tracker used a technique called pupil center corneal reflection (PCCR). The eye tracker would utilize invisible light to illuminate the eye. And capture the reflection from the eye. By analyzing the pattern of the image, the eye tracker can determine the position and rotation of the eye and where the gaze is at. The first purpose of eye tracker is to establish ground truth for the experiment. In order to understand the emotion of the participant during the experiment, a video of screen recording with gaze movement overlay is rendered. (Figure 2) This video is used
Figure 1: Eye Tracker attached to the bottom of the monitor

Figure 2: Gaze indicator rendered on the screen recording video
to understand where the attention of the participants is when certain event happen. The second purpose of the eye tracker is to track the distance between the eye of the participant and the eye tracker itself. Proving the hypothesis of the distance between the eye and eye tracker can be interpreted as posture is one of the goals of this pilot research.

**Audio Recording Device**  A smartphone is placed on the desk under the monitor to record verbal responses from the participant. The purpose of the verbal recording is to establish ground truth. The participant is asked to do think-aloud during the game play session. These would help us identify the emotion of the participant during certain events.

### 4.1.2 Experiment Protocol

Before participants arrive at the lab:

1. Turn on iMotions.

2. Check that if the eye tracker is connected.

3. Welcome and brief participants: “Today you will be playing 3 levels of Candy Crush. Please keep speaking out loud what you feel while playing the game, your voice will be recorded.”

4. Have participants sit in front of the computer.

5. Exam the eye trackers status, make sure everything is well connected.


7. Start the iMotions’ data recording process.

8. Re calibrate eye tracker.
9. Start Candy Crush while researcher verbally indicates start so screen recording and audio recording can be synced afterward.

10. Stop the participants after they finish 3 levels.

11. Stop iMotions from recording.

12. Stop audio recording.

13. Thank the participants.

4.2 Pilot Study Result

Missing eye distance data is being removed and the data is being visualized by in Rstudio with “dygraph” package. The data shows player tend to move closer to during the session and there are peaks and valleys on the graph that represent rapid movements of the the head.(Figure 3)

Combining the think-aloud verbal feedback and screen recording, peaks and valleys in the graph represent emotion changes of the player.

**Small Valley** During 70-75s, 104s, the participant is highly engaged in the game and did something very satisfying. For example, player went on a score strike, so finished the level.

**Small Peak** During 222-243s, 292s, 34s, the participant is trying to solve a puzzle and focusing on the game.

**Huge head movement** Huge head movements represent a huge emotion change on participants part. Huge spike on 196s, 447s, the participant is very happy for the eventful ending when she beat a level, During the huge spike on 313s, the participant is very frustrated by the a microtransaction prompt.
Change pose  During 260s, the participant changed her pose and moved away from the screen because he is confused about the game.

There is a strong correlation between head movement and the enjoyment of the content when players play video games. Especially the rapid head movement that indicate emotion changes. This pilot study proves that eye distance based posture identification can be a good indicator for engagement and there is a good potential on doing quantity engagement assessment based on eye distance.
5 Main Study Methodology

5.1 Rapid Affective Game Response Collection (RAGRC)

Biometrics can provide accurate and objective feedback of players’ emotion for game researchers and designers. With the help of analysing biometrics game researchers and designers can get real-time, objective insight without interrupting players gameplay experience.

5.1.1 Establishing ground truth

Gaining ground truth is one of the biggest challenges for game researchers when biometrics are involved. Since establishing ground truth can require an understanding of context, or get feedback from the players directly. In this research, ground truth will be collected via direct feedback from players since we are planning on a sample size more than a hundred. And setting up ground truth with context manually by the researcher can take a long time and be very challenging. The common methods to gain ground truth include survey, think-aloud, continuous annotation. For this research, we will be using a survey.

Gaming experiences are very hard to distilled into simply good and bad. Even if we category those experiences into good or bad, the reason behind it may vary. That is why we want to collect as many samples as possible, so we can make sure there are more samples in each category of gaming experiences so the category algorithm can properly. For each sample session, the shorter the session, the more likely there are fewer categories of emotion contained in that session.

Thus, we need a way to get sample sessions that are as short as possible.
5.1.2 Solution

The solution are we will be presenting participants with more games than they can play at a time, and told them they can play as many games as possible and stop playing the game as long as they don’t feel like playing the game anymore. In this research, the brief session is being phrased carefully to make sure that the gameplay experience stays authentic to the real gameplay experience. Here is the brief:

“In this exercise, you will play a continuous stream of games on a web browser for 15 minutes. You are welcome to play each game for an as long or short amount of time as you would like. Once you want to stop playing a game and start playing a new one, you can close the game you were playing and open the “New Game” Tab, there will be a link for the new game. There are more games to play than you can play in 15 minutes. I will let you know when the time is up.”

In order to make sure participants don’t stayed too long in a game they like, we mentioned there are more games to play than they could play.

In order to make sure participants are not biased in the quality of the game, we told them they can play each game for an as long or short amount of time as they please. We do tell participants to stop playing the game and find a new one is possible but we believe this can create a lot biased since participants are expected more than one game when they hear “There are more games to play than you can play in 15 minutes.”

In order to make sure participants can switch games easily, we choose web games as stimulus and participants can easily close the current game’s web page and find a link for the new game on the google doc.
In order to prevent game researcher interfere with galvanic skin response, participants are asked to switch to a new game by themselves.

This solution also solved the mixed emotion experience problem, since participants would stop playing the game as soon as they found the game not enjoyable. Makes data of game session collected more prone to have less mixed feeling.

After finished all the game provided or reach the time limit, participants are asked to take a survey on the browser, and they were asked to voice opinions on every game they have played during the session. Participants were asked: "Would you like to play this game again after the experiment?" The answer could be “Yes, I would.”, “No, I wouldn’t” and “I did not play this game during the exercise.” (Figure 4)

If participants answer they would play the game again in the survey, the gameplay session is considered to be pleasant. If not, the gameplay session will be considered as not pleasant.
Figure 4: Survey used in the experiment
5.1.3 The Games Used In The Experiment

For the sake of variable control, all games picked are from the same web-base game provide “Newplaygrounds.com”. Games picked are all arcade game that requires quick response from the players. (Figure 5)

All games are presented in the exact orders for every participant. The second game is a very bad game called “Kill Hitler”, the game received a 3.2 out of 5 score on the website due to major bugs, having a very humble visual presentation and a very simple mechanic. This game do achieved the intentional effect since 14 out of 15 participants find the game enjoyable.

This game is placed in the second place to serve as a reminder for participants that bad game does exist in the list of the games they are playing, and they are free to give up on the game they don’t enjoy.
Figure 5: The website Newplaygrounds and games in it
5.2 Equipment and Experiment Environment

The participant is seated in front of a monitor in a soundproofing cubicle room without any distractions. Researchers are seated next to the participant and give instruction to the participant when needed. An eye tracker is attached to the bottom of the screen.

5.3 Eye Tracker

The eye tracker is the same one used in the pilot study. The eye tracker is used to measure the distance between eye and the eye tracker.

5.4 Galvanic Skin Response Sensor

The GSR sensor is a Bluetooth module that connects with the computer wireless. The player sensor is attached on the participant’s left wrist no matter the participants is left-Handedness or right-Handedness since the control scheme of the games selected for the participant always requires a mouse and keyboard.(Figure 6) A simple pilot study conducted before the main study indicates that the GSR signal noise generated via operating a keyboard is significantly smaller than the noise generated via operating a mouse.

According to the study from Scerbo, Freedman, Raine, Dawson, & Venables(1992)[1], The distal phalanges of the fingers are most likely to have greater galvanic skin response. The GSR electrodes are placed on the participants middle and index finger.(Figure 7)
Figure 6: The GSR sensor module attach to the participants wrist

Figure 7: The electrodes placed on the finger
5.5 Experiment Protocol

This is a joint experiment with another game research project BlockVania, the participant will be playing the game BlockVania first up to 30 minutes, then they are allowed to rest for 10 minutes if they wish so before playing games set up for this experiment. Before participants arrive at the lab:

1. Pull up the survey for BlockVania on the browser.
2. Pull up the survey for this research on the browser.
3. Pull up the “New Game” Google Doc with link of the game in it on the browser.
5. Turn on iMotions.
6. Check that the GSR sensor is powered and can connect.
7. Welcome participants: “Today you are going to be playing a video game. While playing the game I will be measuring your Galvanic Skin Response (or the amount you sweat), and eye movement. We will be doing this through an eye tracker and two skin sensors that are attached to this headband.”
8. Explain the game BlockVania to the participant.
9. Have participants sit in front of the computer.
10. Exam the real-time GSR sensor data and eye trackers status, make sure everything is well connected.
11. Start the iMotions’ data recording process.

12. Re-calibrate eye tracker.

13. Start Game BlockVania from the desktop.

14. After participants finished the game or hit 30 minutes mark.

15. Stop the iMotion from recording.

16. Pull up survey browser. And let participants fill out the survey.

17. Remove the GSR sensor if the participants wish to rest for 10 minutes.

18. Reconnect the GSR sensor to participants after they finished their rest session.

19. Pull up the browser for “New Game” tab.

20. Brief the participants about the games for this experience. “In this exercise, you will play a continuous stream of games on a web browser for 15 minutes. You are welcome to play each game for an as long or short amount of time as you would like. Once you want to stop playing a game and start playing a new one, you can close the game you were playing and open the “New Game” Tab, there will be a link for the new game. There are more games to play than you can play in 15 minutes. I will let you know when the time is up.” Start the iMotions data recording process.

21. Start iMotions recording session.

22. Re-calibrate eye tracker.
23. Participant play the game

24. Stop participants from playing the game after 15-20 minutes of play.

25. Pull up the survey for this research for participants to fill out.

26. Thank the participants
6 Main Study Results

In this section we present the results obtained through the main study in this project, following the protocol described in the previous section.

6.1 Data Overview

There are 107 gameplay sessions collected from 15 participants. The sample is not completely balanced: 36 percent of the gameplay sessions are considered an enjoyable experience by the participants and 64 percent of gameplay session are considered not enjoyable. (Figure 8) The raw galvanic skin response sensor data, eye tracker data, and footage of screen recording is recorded.

In this research, we want to explore some basic metrics first, for GSR
data, we are going to look at the skin conductivity response. And for the eye distance data, we are going to perform peak detection as we did in the pilot study but with a quantitative approach that can be performed by an automatic system.
6.2 Data Prepossessing

6.2.1 Galvanic Skin Response Sensor Data

Ledalab is selected for data prepossessing the galvanic skin response sensor data due to the extraordinary research on galvanic skin response for years and reliable analysis method[6][5].

The galvanic skin response sensor data are noisy due to the hand of the participants constantly interacting with the keyboard (see Figure 9).

In this research, we will be using the adaptive smoothing function in Ledalab[6][5].(Figure 10)

After data is smoothed, a skin conductivity response detection analysis is performed by Ledalab and an “SCR list” is generated. This list will contain the times all skin conductivity responses occurred.

6.2.2 Eye distance data

The eye distance data has a relatively large portion of missing data. Eye blinking is one of the main reason data is lost. In order to make sure the
Figure 10: GSR sensor data smoothed by Ledalab

Figure 11: Peak and valley detected by Quantmod, Red dots indicate valleys and blue lines indicate peaks.

peak and valley of distance detection works properly, we need to make sure there are no missing data. Since we are only interested in peak detection, the data imputation process is rather simple, we are imputing missing data with linear regression based on time. So the peak detection won’t be affected by the imputation method.

We are using an R package “Quantmod”[13] to perform peak and valley detection on the imputed eye tracking data. (Figure 11)
Quantmod is an R package that is designed for quantitative financial modeling and trading but is generally useful for processing time series data. The “findpeaks” and “findvalley” functions are used to detect peaks and valleys in the eye distance data so detection of rapid head movement becomes possible in the quantitative approach.

6.3 Data Analysis

6.3.1 Direct Comparison Between Two Categories Of Gameplay Experiences

Eye distance peak count, eye distance valley count, and skin conductivity response count for each session is scaled with the length of gameplay session since the game lengths may vary between participants (Table 1).

<table>
<thead>
<tr>
<th>Game Experience</th>
<th>Eye distance peak count per second</th>
<th>Eye distance valley count per second</th>
<th>Mean SCR per second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>0.00738544</td>
<td>0.009726711</td>
<td>0.1905814</td>
</tr>
<tr>
<td>Negative</td>
<td>0.00567316</td>
<td>0.00493711</td>
<td>0.1812695</td>
</tr>
</tbody>
</table>

Table 1: Means of two types of game experience’s metric

The average eye distance valley count per second and skin conductivity response count per second between two groups of games are very close to each other. However, the eye distance peak count per second of a positive experience is often larger than the negative game experience. Which indicates that the players tend to lean back more often when they enjoy the game.
6.3.2 Mann-Whitney U Test

Since none of three metrics are normally distributed, a Mann-Whitney U test will be performed in order to evaluate whether there is a statistically significant difference between the biometric data recorded during positive and negative gameplay experiences, respectively.

This test can be used to determine whether two independent samples were selected from populations having the same distribution[1].

According to the result (Table 2) It appears that there is a significant difference of eye distance valley count and eye distance peak count between two game experiences.

<table>
<thead>
<tr>
<th></th>
<th>Eye distance peak count per second</th>
<th>Eye distance valley count per second</th>
<th>Conductivity response count per second</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>968</td>
<td>1023.5</td>
<td>1155</td>
</tr>
<tr>
<td>n1</td>
<td>38</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>n2</td>
<td>69</td>
<td>69</td>
<td>69</td>
</tr>
<tr>
<td>p</td>
<td>0.01969</td>
<td>0.01082</td>
<td>0.3114</td>
</tr>
</tbody>
</table>

Table 2: Mann-Whitney U Test result
6.3.3 Mixed-Modality

The low correlation between skin conductivity response count per second and the experience of gameplay session is expected since a skin conductivity response can represent multiple different feelings. We want to see if multiple metrics can compensate with each other and provide a more accurate insight into the experience.

We tried using the clustering methods k-means[19] and dbscan[9] but the result is not ideal. Both cluster algorithms can’t produce clusters that have distinct game experience in it.

The clusters generated by dbscan can only generate two imbalanced clusters that don’t reflect the game experience we want (Figure 12).

Using elbow method to select the number for clusters with the k-means clustering method, we can produce 3 cluster that seems balanced (Figure 13). But each cluster contains both good and bad game experience dis-

Figure 12: Clusters generated by dbscan
Figure 13: Clusters generated by k-mean method

turbed evenly just like the whole sample size, which we can not use this method to identify the game experience (Table 3).

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Good Game Experience</th>
<th>Bad Game Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 3: Session count in each cluster
7 Discussion

7.1 Findings

7.1.1 Galvanic Skin Response

Even though using the galvanic skin response alone to get insights of experience alone is expected, the skin conductivity response count for both good and bad experience surprisingly similar.

My assumption would be this is created due to the arcade nature of the game, but future research is needed.

7.1.2 Eye Tracking data

The eye tracking data proved to be useful even when being used along since a significant difference of eye movement peak count per minutes is revealed by the Mann-Whitney U Test. Eye trackers are still new to the game research scene and I think the research with eye distance alone can be very promising.

The eye distance peaks(head lean backward quickly) are interpreted as positive feelings or focus in the pilot study, which proves to be the same case in the main study, where the average head peaks per seconds of positive experiences is twice as much as the negative experiences’.

However, Candy Crush and games used in the main study have very heavy arcade elements in it, which requires a lot of attention from the players, the result may vary between different genres of games. Further study
is needed and I believe this is a very promising direction for game and psychology research in general.

7.2 Obstacle to build a mixed modality model

7.2.1 The Numerous Personality And Limited Data Set

This might be one of the reasons why clustering is not working as intended. The sample size is just too small to perform effective analysis. Even if there are 107 samples, there are only 15 participants, this might be the reason behind the failed clustering attempt since each person may react to the same stimulus differently.

7.2.2 GSR

Besides the different reaction from different person discussed above, the GSR signal collecting method also has room for improvement. Even though putting electrodes on the finger proved to be the most effective way, it is not a good measure method for gaming since games involve a lot of hand movement. Collecting GSR data when gaming is a topic worth looking into.

7.2.3 Head Peak Detection

The valley and peaks of raw eye distance data are the most interesting find in this research, the data collecting is simple, however, there is no specialized valley and peak detection algorithm for head movement.
The algorithm may study used is for finance and stock price analysis. A further study for eye distance peak and valley detection will be very helpful.

7.2.4 Ground Truth Not Accurate Enough

In the main study, multiple methods are used to make sure each gameplay session should be as short as possible, however, some gameplay sessions lasted for a long time, which means there is still room for more than one emotion take place in one session.

7.2.5 Data Process Method And Lack of Feature

The biometric data is hard to normalize due to there is more than one participant in the study, each participant sweet differently and have different height, which makes features like the area under the curve unusable.

This makes developing a system that works on most of the population very hard and that reason behind the lack of features in this study.

However, there is also a data analysis method that is not explored in this study due to data constraints, sequences study based on recurrent neural networks.

7.3 Long Short-Term Memory Based Recurrent Neural Network

“Long short-term memory (LSTM) is an artificial recurrent neural network, (RNN) architecture used in the field of deep
7.3.1 Advantage

LSTM is great for time-based pattern reorganization since it will remember events selectively that happened before center time and bring context into consideration, which is a perfect fit for our study goal.

Also, like all neural network, LSTM cannot be influenced by noise easily, given how noisy the GSR signal is, it makes it very particle candidate for game-related analysis.

LSTM can also give feedback in real time, it can be a very powerful tool not only for game researchers but developers who want to make games based on players real-time emotion.

7.3.2 Challenge

In order for neural networks to work properly on most of the population. Training Neural Networks requires a huge amount of diverse data, collecting data would take a very long time. A better data collecting method is needed.
8 Conclusion

In this study, an mixed modality game engagement assessment system is trying to be built with two biometrics features. A classic feature, galvanic skin response, measuring how much people sweat and how excited people are. And a very novice feature, the distance between eye and monitor. This feature is inspired by studies related to postures and attention. A pilot study is conducted, and it confirms that sudden eye distance changes do related to emotion changes.

The study original goal is not completed due to lack of data and limited data processing method. However, other findings were made. The most exciting founding should be the correlation between eye distance and engagement.

The study also discussed what should be done to conduct further study in order to build the target system for game researcher and game designers.
References


[27] Karolien Poels, Yvonne de Kort, and Wijnand Ijsselsteijn. “it is always a lot of fun!”: Exploring dimensions of digital game experience


