Affective reactions to playing digital games

Agnieszka Landowska
Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology
Gdansk, Poland
nailie@eti.pg.gda.pl

Michał R. Wróbel
Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology
Gdansk, Poland
wrobel@eti.pg.gda.pl

Abstract—The paper presents a study of emotional states during a gameplay. An experiment of two-player Tetris game is reported, followed by the analysis of the results - self-reported emotional states as well as physiological signals measurements interpretation. The study reveals the diversity of emotional reactions and concludes, that a representative player's emotional model is hard to define. Instead, an adaptive approach to emotion recognition and individualized classifiers seem to be a better solution for affect-aware digital games.

Keywords: affective computing, digital games, affect-awareness, emotion monitoring, physiological symptoms of emotions

1. INTRODUCTION

Digital games belong to the wide area of entertainment applications. They can strongly influence players’ emotions, even as much as real life events. This is despite the obvious differences between the real world and its digital representation [1]. However, not all affective states induced in players are positive, e.g. horror games stimulate fear and surprise (and relief afterwards). Even though not all gameplay emotions are positive, people keep playing the same game types and sometimes even the same games. At the development stage of a computer game a theoretical model of a representative player is assumed, that consists of both cognitive and mental aspects [2]. However, this approach holds two major drawbacks: (1) each player differs in some way from that averaged model and (2) player’s affective state can change radically within and between sessions. Although emotions are the main reason for playing computer games, there are few studies on which emotions are induced during a gameplay and how they change along the way [3].

This paper presents the design, execution and results of an experiment, aimed at observation of emotions encountered within a multi-session play of two-players Tetris game. The study aims at identification of affective state diversity and change during a gameplay. The experiment hypothesis might be formulated as follows: between-subject diversity of emotional reactions characteristics does not allow to assume a representative player emotional model. The experiment’s design, execution and results are provided in the paper followed by the discussion of the outcomes. The paper is organized as follows: section two provides a short literature review of affect in games research, sections three and four present the experiment design and execution respectively. Sections five and six present the results of the experiment, followed by some discussion in section 7.

The results of the experiment are the first step in construction of affect-aware games, that take into account players’ emotional states. Due to the diversity of gamers’ emotional reactions, adaptive games seem to be better solution for gameplay affective models.
2. RELATED WORK

Traditional digital entertainment does not allow a deeper involvement. In contrast, the emotional experience during a game differs from the reception of films and other media, where emotions can be experienced at most from the third person perspective [1]. Brown and Cairns during their study distinguished three levels of involvement in a virtual game: engagement, engrossment and total immersion. The transition to the latter two levels is possible only when the players' emotions are affected directly by the game [4].

Emotional engagement is required for players to be deeply involved in a game. Events in the game affect the emotions of the players, but also players' reactions in the games are affected by their emotional state. Yannakakis et al. have gone further by proposing the concept of affective loop, according to which games should evolve and adapt to the player emotions in many different ways [5]. Gaming researchers agree on the importance of an affective adaptation of a game to a player emotions as a factor influencing games involvement and effectiveness [6].

Tools developed in the field of affective computing can be used to develop affect-aware games, where many complex emotions can be induced both for entertainment as well as for therapeutic purposes [6]. Therefore, taking into account the study of players emotions can lead to the development of better game projects [5].

Numerous studies have been conducted on the impact of changes in features of the game on the players emotions and their commitment [7]–[9]. Chumbley and Griffiths proved that inducing more negative emotions in relation to positive ones increased the player's frustration and decreased his excitement. In contrast, the strengthening of positive emotions increased the chances of player returning to the game [8]. Qin et al. conducted a study on the level of player immersion depending on the game difficulty [9].

Novak et al. conducted an experiment using the affective loop concept in the popular Snake game. Based on the questionnaire responses about players emotions, the difficulty of the game was adjusted. The experiment results showed the correlation between difficulty adaptation algorithm evaluation and entertainment experience. [10]

Another way to gain information about user emotions is the use of biofeedback – a set of sensors that read the body's physiological responses. Gilleade et al. in their article presented a video game in which the player's heart rate affected the number of threats in the game. Tests showed that even a simple affective loop, using biofeedback sensors, created more pleasant gaming experiences. However experiment also showed a great difference between the players' physiological signals. [11]

In their experiment, Drachen et al. revealed the correlation between heart rate, electrodermal activity and player experience in First-Person Shooter (FPS) games. The player game experience was measured using self-report In-Game Experience Questionnaire (iGEQ). [12]

As stated in Boyle’s et al. review, experiments using biofeedback, conducted so far, were quite simple. None of them used multimodal data inputs, and the use of the individual sensors has many limitations. For example, the heart rate may increase in response to emotional arousal or decrease in response to high levels of engagement. [7]

Conati et al. conducted an experiment that monitored users' physiological responses while playing educational games. Their conclusion was that it was very difficult to assess the users' emotional states in an uncontrolled environment. The level of noise in the biosignals was high due to the fact that sensors were motion sensitive. [13]

There are other studies that confirm the susceptibility of physiological readings to movement artifacts and propose alternative sensors locations [14].
3. EXPERIMENT DESIGN

To verify the research hypothesis that *between-subject diversity of emotional reactions characteristics does not allow to assume a representative player emotional model* an experiment was designed and conducted. In the study of emotions during a gameplay two research methods were employed: a controlled experiment and a questionnaire. Full randomization of subject selection was not possible in a single location, therefore, type II randomization was performed while assigning numbers (play order) within groups. Moreover, the group of convenience was constructed in such a way that it represented some range of possible confounding variables such as age and gender.

The experiment used an emotion monitoring stand that was constructed at Gdańsk University of Technology [14]. The stand was equipped with biometric sensors designed to measure the physiological parameters that are most commonly used in emotion recognition, including: heart rate, skin conductance, respiration, muscle activity, peripheral temperature as well as brain activity [15]–[22]. Heart rate can be obtained from electrocardiographic sensor placed on chest or from blood-volume pulse sensor, placed on finger tips [16]–[18]. In the experiment a fingertip sensor was used. Electromyographic sensors used in emotion recognition are placed on face muscles to track microexpressions and on trapezius muscle (shoulder) because it reflects stress [19], [20]. In the experiment the latter location was chosen. In emotion recognition, typical location of skin conductance sensors is the base of ring and index finger or little and middle finger [21] and the one was chosen in the experiment. Respiration was be measured on thorax, while peripheral temperature on finger tips. Detailed location of sensors in the experiment is provided on Figure 1. Please note, that sensors were placed on a non-dominant hand in order to leave the dominant one for playing. The experiment was followed by a questionnaire to determine subjective assessment of emotions during the gameplay.

The experiment was held in groups of four players. The experiment process consisted of the following steps: (1) information on investigation and signing consent, (2) number assignment within groups, (3) tournament that consisted of matches between each pair of the participants within groups, (3a) attachment of sensors, (3b) measurement sequence (steps 3a and 3b were held for each pair), (4) filling in a questionnaire.

During the experiment, the participants played the two-person version of the popular game Tetris (Figure 2). The aim of the game is to prevent the board from being filled with blocks, by creating horizontal lines without gaps. An open source clone called LTris was used in the experiment.

![Figure 1. Location of sensors in the experiment](image-url)
Tetris has already been used in experiments related to emotions conducted by Lankoski and proved its usefulness. Lankoski presented emotion sequences induced by Tetris structure. At the beginning fear is experienced when a dropping block threatens a goal. Such fear increases arousal and vigilance, which helps to control the game. Next happiness appears if the block is placed correctly or sadness if the action fails. [3]

The game took place between two players. To control the blocks they used the same keyboard, the first player used the arrow keys, the other one used letter keys W, A, S, D. In contrast to the traditional single-player version, an additional element of competition appeared. If during the game the player cleared more than one line of blocks, the lines with random gaps appeared on the opponent’s board. Thus the set of events that could affect the user’s emotional state was extended by two more events.

In order to gather more data LTris program source code has been modified. The extensions allowed to log game events with timestamps to a text file. The following events were logged: starting the game, clearing lines, sending the lines to the opponent, receiving the lines from the opponent, increasing the filling level of the board and ending the game. The data gathered in this way were used to examine whether and how the game events affected the physiological responses of players.

After conducting the experiment (after each tournament), the participants were asked to fill in a questionnaire on emotional states, they encountered during the gameplay. Both positive and negative emotions were included in the questionnaire. Additional questions were asked on:

- how stressful were the game events including losing a match or a set, rising velocity of falling blocks (after the next level was reached), last lines before the end, lines dropped by the opponent player or a mistakenly placed block,

- how positive were the following game events: lines sent to the opponent player, winning a set, a match and a tournament,

- how stressful were the environmental conditions including: video camera recordings, physiological sensors, or audience comments and reactions.
4. EXPERIMENT EXECUTION

The experiment was conducted in 2014. A group of 32 subjects was involved in the experiment, including 11 females and 21 males, with age ranging from 20 to 61 years. Only two participants were left-handed, the majority was right-handed.

Participants were grouped into 8 tournament groups of four. Within a tournament each player played one match with every other participant once (single round-robin schedule), so six matches during each tournament were played. The player who first won two sets, won the match, therefore one game consisted of two or three sets. 48 matches and 108 sets were played in total.

The experimental setting at the emotion monitor stand used the following infrastructure: physiological sensors set, analytical device FlexComp Infiniti by Thought Technology, Canada, that allows to simultaneously sample multiple channels with high frequency (up to 2048 Hz), a computer for stimuli display, a computer with Thought Technology BioGraph Infiniti application for gathering biometric data from the sensors. The physiological sensor set included: skin conductance sensor (0 – 30.0 μS ±5%), electromyography sensor (0-2000μVRMS ±5%), temperature sensor (10 – 45°C ± 1.0°C), as well as blood-volume pulse and respiration sensors that use relative percentage measure (1 – 100% ±5%). All of the sensors were produced by Thought Technology, Canada [22]. An exemplary session recording in Thought Technology application is provided in Figure 3. The left part shows physiological responses (electromyography, skin conductance, respiration, temperature, blood volume pulse and heart rate) of player 1, while the right side of player 2 respectively.

To control the game, the players used their dominant hand, while the sensors were connected to the other hand. The game was held on a single computer and both players used the same keyboard. Figure 4 shows the actual photos taken during the experiment.

As a result of the experiment three sets of data were collected. The first contained data from the device FlexComp Infiniti sampled at a frequency of 256 Hz. In addition to the raw data from the sensors the set contained intermediate and processed data, e.g. heartbeat. The second set contained information about events during the games, recorded by the modified version of LTris program.

Each tournament lasted for about half an hour. Immediately after the tournament recordings were finished, the participants were asked to fill in a questionnaire on affective states they encountered during the gameplay and the experiment.

![Figure 3. Exemplary session recording of two players (one set).](image_url)
5. **QUESTIONNAIRE RESULTS**

After the tournament the participants were asked to evaluate the occurrence of emotional states during the gameplay, using the following scale: 5-most of the time, 4-often, 3-medium, 2-rarely, 1-never. The emotion list included six positive affective adjectives (happy, enthusiastic, satisfied, optimistic, enjoying, content) and six negative ones (unhappy, frustrated, disgusted, disappointed, angry, gloomy).

As reported emotional states were diverse among participants, the data was also grouped with sex, age and left/right-handedness in search for a significant observation, that might allow to formulate subgroup models for representative player.

The mean and standard deviations of the reported emotions for all 32 participants (total) and for male/female subgroups are provided in Table 1 (the significance of means was confirmed by t-test results - all means are statistically significant with p<0.01). Significance in the last column concerns the difference between males and females, which was evaluated with the independent groups Student’s t-test.

Please note, that the most frequent emotional states were positive, as the players were rather amused by the game and the challenge. From the negative emotions set frustrated state was reported more frequently than the other ones, especially among male participants. Less frequent positive states were reported by males, and that result is statistically significant. Female respondents reported more positive states and less negative states, however for the negative ones, the difference with males’ reports is statistically insignificant. Even though there was a difference in the questionnaire results between male and female participants, the intra-group differences were also significant. Based on these results, the representative player models cannot be set for male/female subgroups. The data on reported emotional states were also grouped with age and left/right-handedness, however no statistically significant differences were encountered.

The second part of the questionnaire considered the events, that occurred during the gameplay, and their influence on the level of stress (negative events) and level of happiness (positive events).
<table>
<thead>
<tr>
<th>Emotion</th>
<th>Statistics: mean(standard deviation)</th>
<th>Significance (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Males</td>
</tr>
<tr>
<td>Happy</td>
<td>3.29</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>(1.08)</td>
<td>(0.97)</td>
</tr>
<tr>
<td>Enthusiastic</td>
<td>3.65</td>
<td>3.33</td>
</tr>
<tr>
<td></td>
<td>(1.16)</td>
<td>(1.23)</td>
</tr>
<tr>
<td>Satisfied</td>
<td>3.19</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>(0.97)</td>
<td>(0.98)</td>
</tr>
<tr>
<td>Optimistic</td>
<td>3.71</td>
<td>3.24</td>
</tr>
<tr>
<td></td>
<td>(1.26)</td>
<td>(1.26)</td>
</tr>
<tr>
<td>Enjoying</td>
<td>3.87</td>
<td>3.57</td>
</tr>
<tr>
<td></td>
<td>(1.08)</td>
<td>(1.07)</td>
</tr>
<tr>
<td>Content</td>
<td>3.61</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>(1.16)</td>
<td>(1.06)</td>
</tr>
<tr>
<td>Unhappy</td>
<td>1.61</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>(0.97)</td>
<td>(0.97)</td>
</tr>
<tr>
<td>Frustrated</td>
<td>2.13</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>(1.19)</td>
<td>(1.19)</td>
</tr>
<tr>
<td>Disgusted</td>
<td>1.45</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>(0.80)</td>
<td>(0.87)</td>
</tr>
<tr>
<td>Disappointed</td>
<td>1.94</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>(1.20)</td>
<td>(1.28)</td>
</tr>
<tr>
<td>Angry</td>
<td>1.71</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td>(1.14)</td>
<td>(1.19)</td>
</tr>
<tr>
<td>Gloomy</td>
<td>1.61</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>(1.10)</td>
<td>(1.15)</td>
</tr>
</tbody>
</table>

The question on negative events used the following scale: 5-very stressful, 4-stressful 3-moderate 2-a little stressful, 1-not stressful. The most negative events included (from the most stressful ones, each event is followed with reported stressfulness): mistake in block placement (µ=4.06, ᴰ=0.88), last lines (µ=3.72, ᴰ=1.20), block speed increase (µ=2.97, ᴰ=1.23), extra lines from the opponent (µ=2.94, ᴰ=1.27), losing a set (µ=2.38, ᴰ=1.24). The significance of means was confirmed by t-test results - all means are statistically significant with p<0.01. The reported stressfulness of the negative events had large diversity among participants, which could be observed on the provided histograms in Figure 5. Please note, that most of the events were rated with each value from the 1-5 range by different participants. The participants’ reports were very inconsistent - it seems that people differ in what does stress them and what does not, even during a gameplay.

![Distribution of answers for selected stressful events](image)

Figure 5. Distribution of answers for selected stressful events (5-very stressful, 1- not stressful).
An equivalent question was asked on positive events and the scale was similar to the previous one. The most positive events (from the most positive): winning a set \((\mu=3.93, \sigma=0.95)\), winning a match \((\mu=3.81, \sigma=1.05)\), winning a tournament \((\mu=3.77, \sigma=1.21)\), extra lines sent to the opponent \((\mu=3.03, \sigma=1.33)\). The significance of means was confirmed by t-test results - all means are statistically significant with \(p<0.01\). Positive events were rated more consistently among participants and there was little difference between their positive impact evaluation.

6. **Measurements results**

A variety of physiological data were collected from the FlexComp Infinity analytical device. Experimental design assumed the use of data from four sensors: electromyography, BVP, temperature and skin conductance, as those signals are frequently used in emotion recognition. However, data obtained from the electromyography sensor were too noisy for further analysis. In this experiment the changes in temperature and heart rate were not significant enough to observe the emotional response. Therefore further analysis concentrated on skin conductance data. Skin conductance is correlated with the arousal and reflects both emotional reactions and cognitive activity. It is widely used in both non-technical and technical domains [2].

The analysis began with normalization of data collected from the skin conductance sensor. This pre-processing of data was necessary due to the large difference in signal level between participants. Without this step, comparative analysis of players’ bio signals would not be possible. For the purpose of normalization, the open analytics platform KNIME was used. In the next step, the Origin software was used to generate waveforms on the basis of obtained data. The total number of 216 charts (two charts for each player in 108 sets) was generated.

Visual inspection of the waveforms revealed a significant difference between the galvanic skin response of the participants. As a result of this observation, three groups of players have been distinguished.

The first group includes low-reactive (in terms of changes in skin conductance) participants. Their physiological response to events related to the game or the environment is weak. In Figure 6, player 1 is a prime example of such a low reactive participant. In both sets the waveforms have similar distributions. Large amplitude results from the normalization of data.

The second group, in turn, contains highly reactive players. The physiological responses of these participants exhibit wide variation. In Figure 6, waveforms for player 2 are characterized by frequent variation from the minimum to the maximum.

In the latter group physiological reactions of participants are characterized by the presence of certain peaks. (Figure 6, player 3). It can be assumed that these peaks may be associated with the events occurring in the game.

To verify this assumption it was necessary to compile physiological data with the records of the events from the Tetris game. For this purpose, the authors developed a Perl script which created a number of graphs using Gnuplot. Each graph contains four types of data: the waveform of physiological response, dropping a line event (green), receiving a line from the opponent (red) and the number of lines on the board (orange).

Visual analysis confirmed that for some participants there is a correlation between changes in skin conductance and events in the game. Such correlation was observed in 15 cases out of 32 analyzed, which gives 46.9% of all participants.
Figure 7 presents two exemplary waveforms, where such correlation is visible. On the first waveform, there is a significant physiological response when receiving lines from the opponent (red peak). On the second a noticeable reaction occurs when dropping more than one line, and receiving lines from the opponent. In contrast, Figure 8 shows examples of a highly reactive participant (top) and a low reactive one (lower). In both cases, the game events were not correlated with the changes in physiological response.
Figure 7. Physiological responses on the game's events

Figure 8. Highly and low reactive players' physiological responses
7. **Summary of Results and Discussion**

In the evaluation of affective states during the gameplay two methods were applied: a controlled experiment and a questionnaire. The results can be summarized with the following conclusions:

- during the Tetris gameplay, positive states were reported more frequently,
- females reported positive states more frequently than males, for negative states the results were inconclusive,
- emotions induced by game events were very diverse - the reports contained full range of scale for different participants,
- it was possible to outline more and less stressful events on the average, differences between positive events were small,
- it was possible to differentiate three groups of participants based on physiological signals: highly-reactive, low-reactive and medium-reactive.
- only for the medium-reactive group there was some correlation between game events and changes in the physiological signals.

The study results permit acceptance of the hypothesis that *between-subject diversity of emotional reactions characteristics does not allow to assume a representative player emotional model*. However, we acknowledge that our approach to this study and analysis has some limitations. The main limitations of the study include: subjectivity of self-report and influence of laboratory environment on emotions during the gameplay. Especially, the results for females reporting positive states more frequently than males should be taken with precaution, as the experiment sample was relatively small and this observation would require confirmation in another study.

To eliminate the subjectivity of emotion-related questionnaire, the experiment was held on the physiology tracking stand. However it only partially addressed the issue, as not for all participants the measurements confirmed the questionnaire results.

Laboratory environment is not natural for playing digital games, therefore some impact on the reported and measured emotional states could be expected. This effect is unavoidable in the laboratory experiments, however we have asked about it in a questionnaire. All external conditions (audience, sensors and camera presence) were rated as almost not stressful (on the average from 1.63 up to 1.72 in 1-5 scale). However four participants have reported the camera presence as stressful or very stressful, the same number of participants considered the sensors as stressful (only one participant considered both factors as stressful).

8. **Conclusion**

The study aimed at identification of affective state diversity and change during a gameplay. the results revealed, that it is hard to define a representative player, regarding the emotional states. Due to the large diversity of evoked emotions in reaction to events, it is highly advisable to create adaptive games, that recognize user's emotional states during the gameplay. Moreover, the emotion recognition algorithm within the games should also be adaptive instead of a universal one.

The future works include development of such algorithms and building games, that adapt to user's emotional states.
ACKNOWLEDGMENT

This study was partially supported by Polish-Norwegian Research Programme operated by the National Centre for Research and Development under the Norwegian Financial Mechanism 2009-2014 in the frame of Project Contract no Pol-Nor/210629/51/2013 as well as by DS Programs of the ETI Faculty, Gdańsk University of Technology.

Authors thank all participants of the experiment and researchers from Emotions in Human-Computer Interaction Research Group at the Gdańsk University of Technology (emorg.eu) for their valuable comments to this study.

REFERENCES