Affective Games Provide Controlable Context.
Proposal of an Experimental Framework

Laura Żuchowska and **Krzysztof Kutt** and **Krzysztof Geleta** and **Szymon Bobek** and **Grzegorz J. Nalepa**[1]

Abstract. We propose an experimental framework for Affective Computing based of video games. We developed a set of specially designed mini-games, based of carefully selected game mechanics, to evoke emotions of participants of a larger experiment. We believe, that games provide a controllable yet overall ecological environment for studying emotions. We discuss how we used our mini-games as an important counterpart of classical visual and auditory stimuli. Furthermore, we present a software tool supporting the execution and evaluation of experiments of this kind.

1 INTRODUCTION

Emotions constitute an important context for interpretation of human behavior. Affective computing (AfC) is a field of study devoted to the computer-based analysis, modeling and synthesis of emotions [14]. In our work in this area, we focus on the use of wearable and mobile devices to support the acquisition and interpretation of bodily signals in order to the detect changes of affective states and possibly recognize the corresponding emotional states of subjects. We believe, that the context-aware systems paradigm considered in computer science, should take into the account the affective dimension [11]. Furthermore, the computer models should be personalized, i.e. take into the account individual differences of human behavior, as well as personality traits [8].

One of the principal challenges in the AfC experiments is the actual process to evoke individual emotions for the training and calibration of computer models. In the psychological literature, some of the typical experimental procedures assume the use of standardized visual and auditory stimuli that are supposed to evoke the specific emotions. From our perspective, such an approach is not sufficient as the experimental situation very often does not seem natural to the participant, furthermore it is not personalized. To tackle this challenge, in our work we employ computer games as the source of specific, rich, natural, yet controllable context to evoke emotions [12].

In this paper we present an experimental setup using affective games to evoke emotions of the participants. The principal contributions include: the design of original video games aimed at AfC experiments, a framework for configuration of experiments using such games, putting these two in the context of the BIRAFFE experiments we conducted.

The rest of the paper is organized as follows: In Section 2 we discuss the detailed motivation of our work. Then in Section 3 we describe an experiment in AfC we conducted to acquire data on the individual affective reactions. In this experiment we used a set of affective games we specifically developed for this task, as described

in 4. Furthermore, we realized that in order to provide flexibility of such experiments, we should have a framework supporting the reconfiguration of such experiments for a range of game levels. We developed a prototype of such a framework, as described in Section 5. A short comparison with other solutions is provided in Section 6. In Section 7 we describe the evaluation of our work. We conclude the paper in Section 8.

2 MOTIVATION

Research on emotions requires, on the one hand, a controllable experimental environment to evoke and detect and emotions, on the other, natural conditions for experiments in order to minimize a possible discomfort for the participants. Video games seem to be a good trade-off between these two extreme requirements. Games allow to control the appearing stimuli and log everything that happened, especially the reaction times, Moreover, the environment is rich in stimuli and allows for user interaction with objects, including emotionally related interaction framed in the so-called Affective Loop [12].

"Regular" games, available on the market, do not meet the requirements of the experimental environment. First of all, they provide a (too) rich environment in which the player may do (too) many things. In such an environment, a very large sample size is needed to get the right statistical power to draw conclusions, which makes experiments difficult to conduct. Also, the use of machine learning methods will not be trivial, as there are many variables in such case, some of which will only be disruptive noise. Secondly, "regular" games do not allow for the evaluation of emotions too often. The player is constantly engaged in the game and interrupting it to complete the questionnaire will reduce the immersion of the game.

These issues have been observed in our previous experiments [11] including the BIRAFFE1 experiment [8]. To address them, a set of mini games, with restricted experimental conditions, was created. Each of them is built up on a very limited set of stimuli, with the aim of evoking a limited set of emotional reactions. The following sections describe an experiment called BIRAFFE2 (see Section 3) in which three such games were used (see Section 4).

The BIRAFFE2 experiment has led to observation of further issues that need to be addressed when conducting game experiments. In particular, attention has been drawn to the fact that all mini games should generate event logs in a uniform format to avoid additional pre-processing steps when analysing the collected data. It is equally important to implement questionnaires directly in the games, at the end of each mini game. Filling out the questionnaires at the end of the gaming session makes the impressions fuzzy and the self-description may not be accurate enough.

Therefore, in parallel with the BIRAFFE experiments, a dedicated

framework was developed to automate the preparation of game-based affective experiments. It allows to generate an experiment template with questionnaires between different levels and provides a database-based logging interface. A detailed description of the framework can be found in Section 5.

3   THE BIRAFFE2 EXPERIMENT

The BIRAFFE2 study included 103 participants (33% female) between 18 and 26 (M = 21.63, SD = 1.32), recruited among students of the Artificial Intelligence Basics course at AGH University of Science and Technology, Kraków, Poland and their friends.

It is a revised version of a previous experiment called BIRAFFE1 (**Bio-R**eactions **a**nd **F**aces **f**or **E**motion-based Personalization) described in [8].The aim of the study was to collect physiological data paired with behavioral data, which can then be used to develop models for prediction of emotions.

Behavioral data were twofold: from the part in which the subjects played three games (for details see Section 4) and from the classical experiment, in which sound and visual stimuli (from IADS [2] and IAPS [9] datasets respectively) were presented and then subjects were asked to assess what emotions they evoked. Specifically, the stimuli was presented for 6 seconds, what was followed by 6 seconds for affective rating with the use of custom widget with 2-dimensional space (valence and arousal). The whole behavioral data was collected as a set of logs in comma-separated (CSV) files.

Physiological signals, Electrocardiogram (ECG) and Electrodermal activity (EDA), were gathered using BITalino (r)evolution kit, as it is the most promising of cheap mobile hardware platforms (for comparison see [7]). Besides ECG and EDA, during the experiment also the following signals were collected: accelerometer and gyroscope from gamepad, facial images taken by webcam (every 250 milliseconds), screencast of the whole game session.

The whole protocol consisted of several phases:

1. NEO-FFI paper-and-pen questionnaire [15] for personality measurement (approx. 10 minutes),
2. Physiological devices setup (approx. 2 minutes),
3. Baseline signals recording (1 minute),
4. Instructions and training (approx. 5 minutes),
5. First part of stimuli presentation and rating (17.5 minutes),
6. Games session (up to 15 minutes in total),
7. Second part of stimuli presentation and rating (17.5 minutes),
8. Three paper-and-pen GEQ questionnaires [5] (one for each game) and gaming experience questionnaire (approx. 10 minutes).

The whole protocol lasts up to 75 minutes. Steps 3-7 were done on a PC. All of them were controlled by the Python 3.8 with the use of PsychoPy 3.2.4 library [13]. Participant interacted with the procedure only with a gamepad.

4   EVOKING EMOTIONS WITH AFFECTIVE GAMES

In order to support the game sessions of the experiment, three specific affective mini-games were created [16]. The aim for all the games was to create an immense amount of emotions in a short time. The main obstacle was the inability to create an intriguing story, therefore the whole section of narrative elements was discarded. The only way of building an affective project was to make different sets of games with a variety of mechanics and audiovisuals.

The simplest solution to create an emotional-changing environment was to revolve around the overall difficulty of games. While making the neutral, peaceful stage can relieve stress for the player, the loud and hard level can intensify the rage and increase heartbeat. Therefore, three genres have been selected: roguelike, platform, and maze. The first level is balanced to be an easy stage, supposed to develop energetic, happy emotions. On the contrary, the second level is extremely hard to beat, filled with traps, to give the sensation of unjust and fury. This juxtaposition is important to the study, given the sudden change. Last phase is neutral, without any emotion-boosting elements, it exists to check the player's decision-making, behavior and bodily changes due to previous irritation. Additionally, a proper collection of game patterns was implemented. For every stage, depending on what emotions it should boost, from that collection separate elements were chosen.

Figure 1.   Stage 1: an example screen of the game

Stage one contains elements such as score tracking, weapons, enemies and looting. The finishing condition is elimination of all antagonists – no stress-inducting time limit was implemented. The difficulty in this level was balanced by setting the damage per second of the protagonist much higher than the one of the antagonist. While players can shoot up to 5 projectiles per second, enemies can shoot only one attack per second. Moreover, the speed of player's projectiles is 2.5 times higher for the default weapon. An additional blaster was placed on a map, giving the possibility for the user to eliminate the enemies even easier. Furthermore, in the case the subject is not used to playing games, health points can be increased by picking up heart-shaped objects. In order to unleash more fun and any form of achievement-getting sensation, a score tracker is incrementing when picking up money bags from the floor or from the killed enemy (the amount of bags dropped from antagonist is random).

In the platform game (stage 2) traps and time limit were implemented. Both of them are crucial in order to imply stress and rage. Until the end of the game, the player has to go through the whole level. However when the player dies, he respawns in the last checkpoint – a yellow flag with letter 'C'; when touched, a happy, although very distorted sound is played. There are two possible ways for the protagonist to lose: falling down off the stage, or stepping into a spike trap. Considering the fact that this level is supposed to be insanely hard to get through, two additional traps were implemented to basic blocks. The first type is an invisible block – before the protagonist collides with them, they are not to be seen in any way by the user. If the player dies after triggering the visibility, it is once again set to in-

visible. Similar mechanics is once again used for next type of traps – falling blocks. Once the collision with the user happens, blocks start to fall down.

Figure 2. Stage 2: falling block before trigger

Figure 3. Stage 2: falling block after trigger

Figure 4. Stage 2: invisible block before trigger

Figure 5. Stage 2: invisible block after trigger

For the last game in stage 3 memorizing the way through a maze is the only important part. No time nor score tracking is implemented. Visuals are very simple, no distracting elements were added. The choices made by the player are saved into logs, which will be discussed later on.

Size and shape of colliders were also adjusted to the game genre and difficulty intended. For the first scene, the collider for the protagonist is smaller than his real model. It removes the feeling of being hit before the projectile hits the player. On the contrary, in the second game colliders are too big. Player can get hit by a trap before he touches it with a model. This decision was made to enhance the irritation and the feeling of unjust. For the last level, colliders were adjusted to not hit the walls too often, so the movement will be pleasant and smooth. Another intentional difference in stage two from others is the protagonist's movement. It was designed similarly to the jumping mechanics, although it doesn't stop at a certain speed – the player's model is constantly given acceleration. This is a perfect example of poorly made mechanics, which are incredibly hard to control.

Figure 6. Stage 3: protagonist colliders for different stages.

To boost the affective part of gameplay, sounds provided by *NIMH Center for the Study of Emotion and Attention* [2] were added to the every stage. They were proven to change the state of user's emotion by their degree of affectiveness. This level was separated into two values – intensity of the feeling (arousal) and pleasantness of a sound. Depending on these two values, proper sounds were chosen and included in the games.

Sound	Pleasure	Arousal
Puppy	2.88	4.91
Bees	2.16	7.03
Vomit	2.08	6.59
Babies cry	2.04	6.87
Baby cry	2.75	6.51
Scream	2.05	8.16
Child abuse	1.57	7.27
Applause	7.32	5.55
Rollercoaster	6.94	7.54
Colonial music	6.53	5.84
Bugle	6.32	6.35
Rock n roll	7.90	6.85
Funk music	6.94	5.87

Table 1. Affective sounds used in study

Furthermore, music themes and in-game sounds were recorded. The design was created with a view to expected emotions. First game's theme consists of electronic/rock music, sounds of picking items are clear, echo has been added to each sound. To keep the second level unbalanced and irritating, time signature for background music was disturbed – the last eighth note was erased. This gives an unsettling feeling, like someone has been playing off tempo. Additionally, each time the player dies increases the pitch and distortion effects for the background theme. Protagonist has a high-pitched voice, which gets more infuriating with every death. The sound of winning (which is hard to achieve, given the difficulty of the game) has a very disappointing and unsatisfying tone. Last level has a pleasant theme, edited to sound like old arcade, 8-bit music.

In order to get as much as possible from single gameplay about the state of players' emotions, additionally to their bodily functions, a proper context-gathering mechanism is required. It is implemented as a set of different event logs that are saved for each stage. Some information is constantly saved, no matter the level – the data about current player's position, ID and timestamp of an affective sound played in the background. For stage one, events such as killing an enemy, death, the amount of all objects picked up, current state of health and points are saved with the proper timestamp. Additionally, the amount of projectiles shot and their accuracy is recorded – this gives more insight on the aggressiveness and gaming experience. There are no enemies and pickable items in the second stage, therefore the distortion rate of music, number of deaths and the data about traps triggered is saved for every iteration. In the last stage, the amount of dead ends encountered and the data about going off the correct path is being saved.

All of the games were developed using the Unity Engine. It is a powerful environment, with tons of possibilities. One of many features used are previously mentioned colliders. The engine contains a variety of collider shapes, components and traits. For instance, Box Colliders were used not only as physical objects, but also as triggers in rooms for the first stage, logging in the third level etc. Animations in all games were handled through the Animator Controller feature. Another remarkable example for the possible power of Unity Engine is Camera - just a simple change in view can drastically change the
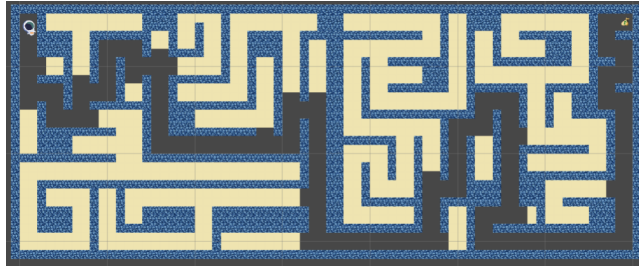
Figure 7. Illustration on how the data about wrong path is saved

Figure 8. 'Feedback' option available in menu

Figure 9. Plugin's use case diagram

perception of the player. While first-person cameras can increase the immersion with the protagonist, third-person cameras can give an insight on what's going on around the character, escalating the feeling of stress for the subject. Moreover, sounds and music have separate components - Audio Listener and Audio Source. Both of them have a simple mixer, included inside the engine. Those components create impressive opportunities for manipulation of emotion. The simplest example used in the game is the distortion attribute, for each death in the second stage.

5 FRAMEWORK FOR GAME CONFIGURATION IN EXPERIMENTS

In order to get information from subjects about their feelings towards the games, the GEQ questionnaire was used [5]. This survey consists of three parts: The Core Questionnaire, The Social Presence Module and The Post-game Module. All of them contain important information about different sections of study. All of them involve questions about feelings, with a range of possible answers from 0 to 4. Zero means 'not at all', one means 'slightly', two is 'moderately', three is 'fairly' and the last, four is 'extremely'. First part of the survey has 33 questions about emotions and sensations felt during the game, for example: 'I was good at it' and 'I felt frustrated'. The Social Presence Module contains 17 questions, however it should only be taken when any form of social interactions were taken in game, whether it's another person or a simple non-playable character interaction. The last section involves 17 questions about the overall feeling of a subject after the game has been played, for instance: 'I felt satisfied' and 'I found it a waste of time'.

To make this questionnaire a part of study, and also to provide a unified context-logging mechanism, a software framework has been written. It is responsible for starting all mini-games and preparing the survey after each game. To install the plugin you need to copy .dll files and prefabs into Unity project. After restarting the editor, you should see "Feedback" menu in the menu bar and the configuration file in "/Resources" directory. In order to start using the plugin, you need to create an SQL database with tables for each survey form you want to include in your game. By default the plugin saves answers as integers, so each question should have a separate column of this type.

Everything is connected through a proper configuration file. It is required to set correct database provider and connection string. After that, the model classes can be generated by choosing "Generate model classes" button from "Feedback" menu. Pressing "Create survey form" button will open wizar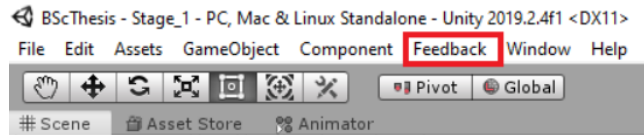d that allows to choose dif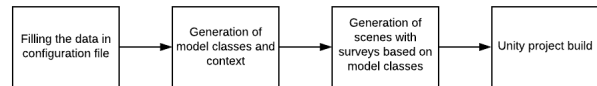ferent types of questions (radio buttons, slider or dropdown). The plugin will create new scene with questions based on the table in database. After that, a request to manually add generated script-handling data persistence to an empty object in the scene will pop up. At the end, there is a possibility to change the text and position of questions in the scene. After building the project, a game is started and the survey pops up next. When all questions are answered, the framework sends all data to the local SQL database.

Figure 10. Survey example

This framework has a very high potential for further studies. Firstly, it gives an opportunity to create a multi-platform study. More computers would be available to use for a study. Furthermore, a mobile version could be implemented. This way, even more subjects would've taken part in a study. Another possible future usage is the adjustment of level difficulty for every game, dependent on answers in the survey. This would increase the affective part of study, as personal change in games would take place.

6 RELATED WORK

There are quite a few different frameworks for affective research. On the one hand, one can point out the tools used to build classical psychological experiments, like PsychoPy [13], OpenSeasame [10],

or E-Prime[2] and on the other hand, the tools used for affective experiments with games, e.g. FILTWAM [1], iHEARu-PLAY [4], or emoCook [3].

The tools in the first group offer various widgets for collecting information from users, making it possible to transfer virtually any paper questionnaire to the electronic version. However, they do not allow one to control a stimulus-rich game environment. This problem has been addressed in the second group of tools, where affective interaction is carried out in games (e.g. educational game "emoCook"). Nevertheless, these solutions are prepared for specific applications and do not provide a general solution for affective experiments.

The framework described in this paper combines the advantages of these two groups. It both allows for the use of games as a research environment and is a general solution, allowing for the inclusion of any games (written in Unity) and any questionnaires (the application is not limited to the GEQ described in the article).

7 EVALUATION

The motivation to introduce a few short mini-games was better control over the emotions evoked during the experiment. The assumption was that each game aim is to evoke specific emotions using a small number of stimuli. These assumptions were confirmed by the results of the GEQ questionnaire.

Revised list of GEQ factors [6][3] was used for analysis. A series of one-way ANOVAs was conducted to evaluate the differences between games. Post-hoc comparisons were done using the Tukey HSD test. Analysis was performed in Python with scipy[4] and statsmodels[5] libraries.

The strongest effects can be observed for the second level, which should give the sensation of unjust and fury. It was connected with significantly higher *Negativity* ($M = 2.85$), significantly lower *Positive Affect* ($M = 1.14$) and significantly lower *Competence*[6] ($M = 0.86$) than the two other stages (*Negativity*: $M = 1.11$ and $M = 0.70$, *Positive Affect*: $M = 2.53$ and $M = 2.49$, *Competence*: $M = 2.42$ and $M = 2.78$, for Stage 1 and Stage 3 respectively).

Stage 1, designed as an easy stage connected with positive emotions, and Stage 3, designed as emotionally neutral, were both evaluated as the ones with the higher *Positive Affect* (there were no significant differences between them). Neutrality of the third level is revealed with the lowest *Negativity* ($M = 0.70$; significantly lower than the first level, $M = 1.11$).

Flow, indicating whether or not players have lost control of their time in the game, was significantly lowest in the third level ($M = 1.35$) than the other two levels ($M = 1.99$ and $M = 2.02$), indicating that for this factor the most important is the fact that emotions are evoked, no matter whether they are positive or negative. Finally, *Immersion*, the subjective connection to the game, was low for all levels ($M = 1.75$, $M = 1.23$, $M = 1.66$ for levels 1-3 respectively), which is also consistent with the assumptions. The games were too short for the players to get fully involved.

We tested the framework on several platforms including Windows, Linux and mobile platforms running Android operating systems. It ran correctly on all of the platforms[7] proving its portability between most popular operating systems. We also tested it with different databases including remote MySQL databases and SQLite database for Android systems, where in both cases it worked correctly. While experiments presented in this paper did not use the framework, they will be used by us as a baseline for future evaluation of the framework.

8 FUTURE WORK AND SUMMARY

In the paper we presented our recent work conducted as a part of the BIRAFFE2 experiment in Affective Computing. As a novel part of the experiment we developed three specially designed mini-games, based of carefully selected game mechanics. We believe, that games provide a controllable yet overall ecological environment for studying emotions. We used these games as an important counterpart of classical visual and auditory stimuli during the experiment to evoke emotions of participants. Moreover, we presented a software tool, with a built-in context-logging mechanism, supporting the execution, automation and evaluation of experiments of this kind.

In the future, we would like to develop our work in several directions. First of all, based on the analysis of the results of the experiment, we will continue the development of new games with improved mechanics to fine tune the evocation of emotions. Ultimately, we expect games will help us in developing computer-based personalized models of emotions to be used in different applications. Furthermore, based on the future findings, we would like to study the aspects of emotional adaptation and personalization in games using the machine learning methods. Finally, our current setup is ready to be used not just in desktop games, but also on mobile devices. We will explore this direction, as mobile games not only constitute a very important market for games, but also offer new opportunities for interaction.

REFERENCES

[1] Kiavash Bahreini, Rob Nadolski, and Wim Westera, 'FILTWAM - A framework for online affective computing in serious games', in *Fourth International Conference on Games and Virtual Worlds for Serious Applications, VS-GAMES 2012, Genoa, Italy, October 29-31, 2012*, eds., Alessandro De Gloria and Sara de Freitas, volume 15 of *Procedia Computer Science*, pp. 45–52. Elsevier, (2012).

[2] Margaret M. Bradley and Peter J. Lang, 'The international affective digitized sounds (2nd edition; iads-2): Affective ratings of sounds and instruction manual. technical report B-3', Technical report, University of Florida, Gainsville, FL, (2007).

[3] Jose Maria Garcia-Garcia, Victor M. Ruiz Penichet, María Dolores Lozano, Juan Enrique Garrido, and Effie Lai-Chong Law, 'Multimodal affective computing to enhance the user experience of educational software applications', *Mobile Information Systems*, **2018**, 8751426:1–8751426:10, (2018).

[4] Simone Hantke, Florian Eyben, Tobias Appel, and Björn W. Schuller, 'ihearu-play: Introducing a game for crowdsourced data collection for affective computing', in *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015, Xi'an, China, September 21-24, 2015*, pp. 891–897. IEEE Computer Society, (2015).

[5] Wijnand A. IJsselsteijn, Yvonne A. W. de Kort, and Karolien Poels, *The Game Experience Questionnaire*, Technische Universiteit Eindhoven, 2013.

[6] Daniel M. Johnson, M. John Gardner, and Ryan Perry, 'Validation of two game experience scales: The player experience of need satisfaction (PENS) and game experience questionnaire (GEQ)', *Int. J. Hum. Comput. Stud.*, **118**, 38–46, (2018).

[7] Krzysztof Kutt, Wojciech Binek, Piotr Misiak, Grzegorz J. Nalepa, and Szymon Bobek, 'Towards the development of sensor platform for processing physiological data from wearable sensors', in *Artificial Intelligence and Soft Computing - 17th International Conference, ICAISC*

---

[2] See: https://pstnet.com/products/e-prime/.
[3] The values are ranging from 0 (not at all) to 4 (extremely) for each factor.
[4] See: https://www.scipy.org/.
[5] See: https://www.statsmodels.org/.
[6] *Competence* reflects how well players judged their own performance against the game's goals.
[7] The only requirement is to use Unity build 2019.2.19f1

2018, Zakopane, Poland, June 3-7, 2018, Proceedings, Part II, pp. 168–178, (2018).

[8] Krzysztof Kutt, Dominika Drążyk, Paweł Jemioło, Szymon Bobek, Barbara Giżycka, Víctor Rodríguez Fernández, and Grzegorz J. Nalepa, 'BIRAFFE: Bio-reactions and faces for emotion-based personalization', in *AfCAI 2019: Workshop on Affective Computing and Context Awareness in Ambient Intelligence*, volume 2609 of *CEUR Workshop Proceedings*. CEUR-WS.org, (2020).

[9] Peter J. Lang, Margaret M. Bradley, and B. N. Cuthbert, 'International affective picture system (iaps): Affective ratings of pictures and instruction manual. technical report B-3', Technical report, The Center for Research in Psychophysiology, University of Florida, Gainsville, FL, (2008).

[10] Sebastiaan Mathôt, Daniel Schreij, and Jan Theeuwes, 'OpenSesame: An open-source, graphical experiment builder for the social sciences', *Behavior Research Methods*, **44**(2), 314–324, (2012).

[11] Grzegorz J. Nalepa, Krzysztof Kutt, and Szymon Bobek, 'Mobile platform for affective context-aware systems', *Future Generation Computer Systems*, **92**, 490–503, (mar 2019).

[12] Grzegorz J. Nalepa, Krzysztof Kutt, Barbara Giżycka, Paweł Jemioło, and Szymon Bobek, 'Analysis and use of the emotional context with wearable devices for games and intelligent assistants', *Sensors*, **19**(11), 2509, (2019).

[13] Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv, 'Psychopy2: Experiments in behavior made easy', *Behavior Research Methods*, **51**(1), 195–203, (2019).

[14] Rosalind W. Picard, *Affective Computing*, MIT Press, Cambridge, MA, 1997.

[15] Bogdan Zawadzki, Jan Strelau, Piotr Szczepaniak, and Magdalena Śliwińska, *Inwentarz osobowości NEO-FFI Costy i McCrae. Polska adaptacja*, Pracownia Testów Psychologicznych, Warszawa, 1998.

[16] Laura Żuchowska, *Game Design with Unity for Affective Games*, BSc thesis, AGH University of Science and Technology, 2020. Supervisor: G.J. Nalepa.