The BIRAFFE2 Experiment.
Study in Bio-Reactions and Faces for Emotion-based Personalization for AI Systems

Krzysztof Kutt and **Dominika Drążyk**[1] and **Maciej Szelążek**[2] and **Szymon Bobek** and **Grzegorz J. Nalepa**[3]

Abstract. The paper describes BIRAFFE2 data set, which is a result of an affective computing experiment conducted between 2019 and 2020, that aimed to develop computer models for classification and recognition of emotion. Such work is important to develop new methods of natural Human-AI interaction. As we believe that models of emotion should be personalized by design, we present an unified paradigm allowing to capture emotional responses of different persons, taking individual personality differences into account. We combine classical psychological paradigms of emotional response collection with the newer approach, based on the observation of the computer game player. By capturing ones psycho-physiological reactions (ECG, EDA signal recording), mimic expressions (facial emotion recognition), subjective valence-arousal balance ratings (widget ratings) and gameplay progression (accelerometer and screencast recording), we provide a framework that can be easily used and developed for the purpose of the machine learning methods.

1 INTRODUCTION

Affective Computing (AfC) [16] – an interdisciplinary field of study regarding human emotions – is to large extent built upon the assumption that we are able to precisely detect, label and manipulate emotional responses of agents. Therefore the proper understanding and modeling of this complex phenomena [1, 3], as well as maintaining ingenious experimental setup to do so, is a crucial determinant of success in this field. The setup should require conditions where human subjects are exposed to specific emotion evoking stimuli, and furthermore their reactions are somehow measured.

In our work we assume that the measurement of bodily reactions to stimuli can serve as proper foundation for emotion recognition (James-Lange approach [7]). We also use a representation of affective data which is common in many experiments in psychology and human-computer interaction, i.e. the two dimensional Valence/Arousal space. We aim to continue a longer effort in building enhanced AfC models, previously presented on the AfCAI workshop, as well as later on in [13]. To make a measurement context as ecological as possible, we use wearable devices as well as other sensors which are possibly non-intrusive to the subjects. We also assume that the data processing should be possible offhand – with the

use of devices that users have with them, e.g. mobile phones. Our recent results in this regards are summarized in [14].

The original contribution of the paper is the report on a an affective computing experiment conducted between 2019 and 2020, that aimed to develop computer models for classification and recognition of emotion. The experiment resulted in the creation of publically developed dataset called BIRAFFE2. We believe that this work might be an important step to develop new methods of natural Human-AI interaction. The rest of the paper is organized as follows. In Section 2 we inroduce the motivation for our work and the context of the previous experiment in this area. Then in Section 3 we discuss our experimental methodology. The structure of the resulting dataset is described in Section 4. The paper ends with a summary and plans for future work in Section 5.

2 MOTIVATION

In order to fully exploit advantages of user's emotion detection modules, whether in games or in apps, the recognition and classification of emotional states must be personalised. By 'personalised' we mean adjusted to ones personality traits and customs. To implement that kind of tailoring, the appropriate variety of information about the user must be collected. In order to verify these assumptions and to gather data for further development of the framework (for summary of the history of our approach see: [13, 14]) we constructed a combined experimental paradigm. First, in the "classical approach" experiment, we presented stimuli to the subjects and collected their answers: both with questionnaires and physiological signals measurement. Secondly, within the "ecological approach", we embedded the reaction measurement in specific context of simple computer games. When a player is loosing, and the system detect the increased intensity of the player's reaction, it is crucial for the model to know the context of such change – only by the contextual information coming from the game progression, we are able to tell whether the intensification was a collateral of anger or joy. Knowing its contextual origin, we can also easily prevent these specific changes from happening again, or on the contrary – elicit detected and labelled emotions once again. As such, games offer a perfect opportunity for putting the human in the feedback loop with the computer system (also called *affective loop*) [5].

The paradigm presented in this paper is the continuation of our previous work on the BIRAFFE1 experiment presented in [10]. In the current experiment we extended the range of the measured responses by the introduction of accelerometer from game pad. Furthermore, we developed new custom computer games dedicated to the elicita-

---

[1] Jagiellonian University, Poland, email: krzysztof.kutt@uj.edu.pl, dominika.a.drazyk@gmail.com

[2] AGH University of Science and Technology, Poland, email: maciej.szelazek@agh.edu.pl

[3] Jagiellonian University, Poland, email: szymon.bobek@uj.edu.pl, grzegorz.j.nalepa@uj.edu.pl

tion of player emotions. Finally, we improved the affective widgets capturing the feedback from the participants.

3 METHODOLOGY

3.1 Participants

103 participants (33% female) between 18 and 26 ($M = 21.63$, $SD = 1.32$) took part in the study. Information about recruitment was made available to students of the Artificial Intelligence Basics course at AGH University of Science and Technology, Kraków, Poland. Although participation was not an obligatory part of the course, one could get bonus points for a personal participation or invitation of friends.

3.2 Questionnaires

First, the paper-and-pen Polish adaptation [18] of the NEO Five Factor Inventory [4] consisting of 60 self-descriptive statements evaluated on 1-5 scale (1 – strongly disagree; 5 – strongly agree) was used to measure the Big Five personality traits.

Second, our own paper-and-pen Polish translation of The Game Experience Questionnaire (GEQ) Core Module [6] was used to measure players' feelings during the game session. The module consists of 33 items, ranked on 0-4 scale (0 – not at all; 4 – extremely). Items were arranged in seven components in original version: Competence, Sensory and Imaginative Immersion, Flow, Tension/Annoyance, Challenge, Negative affect and Positive affect. The questionnaire has been used in many game studies [8, 12]. However, the 7-factor structure has not been confirmed by anyone. In [8] revised version (GEQ-R) was proposed. Tension/Annoyance, Challenge and Negative affect were merged into Negativity, leading to 5-components solution.

Finally, our own simple questionnaire was used to measure gaming experience. It consists of two questions: (1) "Over the past year I have played computer / mobile / video games:" (2) "In the period of my most intense interest in computer / mobile / video games, I played:". Both were answered by selecting one of the five possible answers: (a) daily or almost daily, (b) several times a week, (c) several times a month, (d) several times a year, (e) not at all. There was also a space for leaving comments on the games and experimental procedure.

3.3 Stimuli Selection

Standardized emotionally-evocative images and sounds from IAPS [11] and IADS [2] datasets were used as stimuli, each characterized by its coordinates in the Valence-Arousal space. The analysis of IADS and IAPS scores revealed the following trend: arousal score increases as the valence score strives for it's positive or negative extreme (Figure 1).

For the purpose of the experiment, we divided the stimuli into three groups according to their arousal and valence index: + (positive valence and high arousal), 0 (neutral valence and medium arousal), – (negative valence and high arousal). Erotic stimuli were excluded from the blind selection, due to the risk of creating weird or disgusting combinations (e.g. picture of kid or snake is paired with the erotic sound), not intended by the aim of this study.

The stimuli set for each participant was generated by random sampling without replacement and formed nine conditions, each consisting of 13 stimuli:
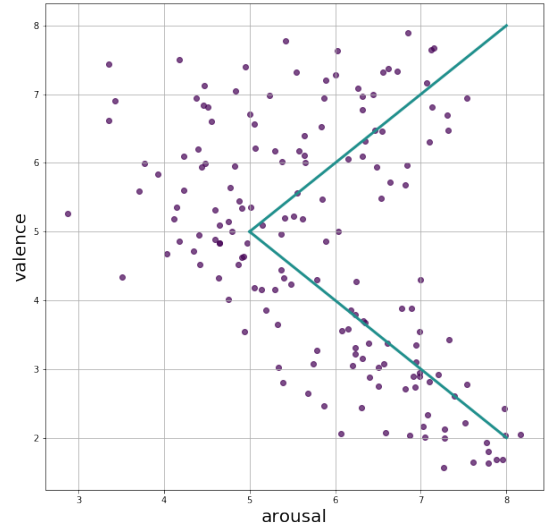

Figure 1. Trends in the IADS stimuli ratings.

- three consisting only of pictures: *p+*, *p0*, *p–*,
- three consisting only of sounds: *s+*, *s0*, *s–*,
- three, where pictures were paired with sounds: + picture with + sound (*p+s+*), 0 picture with 0 sound (*p0s0*), – picture with – sound (*p–s–*).

Conditions were mixed during the presentation, which was divided into two sessions (17.5 min each) and separated by the game session.

3.4 Emotion Evaluation Widget

Emotional assessment was carried out using the *Valence-arousal faces* widget controlled by a left joystick on a gamepad. Widget was adapted from our previous experiment [10], with the following improvements:

- Emoticons placed as hints were moved outside the selection area. Also, the border of the selection area was introduced. In the previous version the subjects often chose the location of the smiley as their answer. Now, there is no possibility to put selection marker on them.
- The selection marker changes color to indicate that there is only half a second left for the rating. In the previous version there was no information about the remaining time.
- The returned valence and arousal scores are now within a range of $[1, 9]$, so they are within the same range as the assessments in IAPS and IADS. In the previous version, they were in the $[-1, 1]$ range, which required conversion of values before the analysis started.

To compare current and previous version of *Valence-arousal faces* widget see Figure 2, *Valence-arousal faces* widgets regard studies (in Polish). X axis has labels "negative", "neutral", "positive", while Y axis has labels: "high arousal" and "low arousal". Pictures presented with a negative filter.

3.5 Games

Three affective games developed by our team [19] were used during the study. All of them were controlled by a gamepad and produced
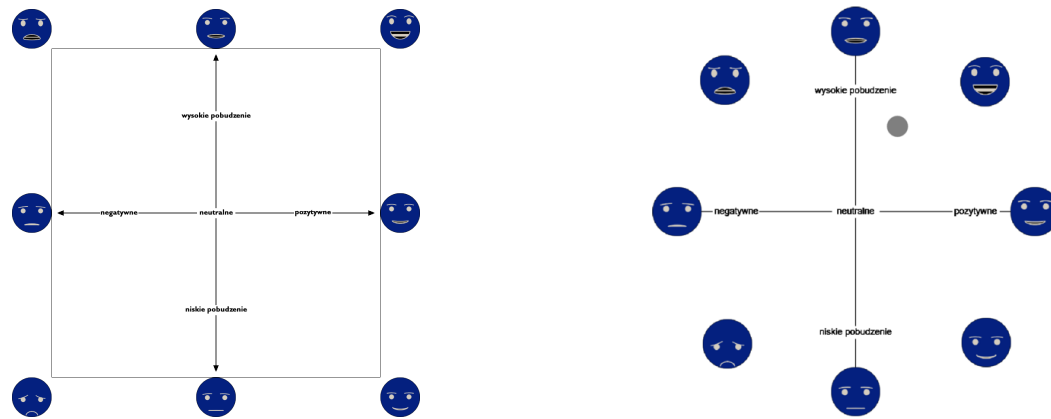
Figure 2. Current and previous [10] versions of the *Valence-arousal faces* widget.

game log CSV files. They have been designed with the emphasis on differentiating the levels of difficulty:

- *Room of the Ghosts*: The goal: pass through a series of rooms and defeat the arriving ghosts (see Figure 3). Difficulty: very easy, achieved by the following implementations: collider for protagonist is smaller than his real model – it removes the feeling of being hit before the projectile hits the player; the protagonist's weapon that can shoot more often and faster than the opponents' weapons.
- *Jump!*: The goal: reach the end of the path by jumping on the platforms and avoiding obstacles (see Figure 4). Difficulty: hard and frustrating, achieved by the following implementations: colliders are too big – player can get hit by trap before s/he touches it with the model; movement is clunky, and there are several traps, i.e. invisible blocks, which increases the confusion and irritation in player; each time the protagonist dies, the background music is getting less pleasant (the pitch and distort levels of music playing in background increases by 0.07).
- *Labyrinth*: The goal: walk the protagonist through the labyrinth (see Figure 5). Difficulty: optimal, achieved by the following implementations: the colliders have been adjusted to not hit the walls too often and to make the movement smooth; the protagonist control is natural and predictable.
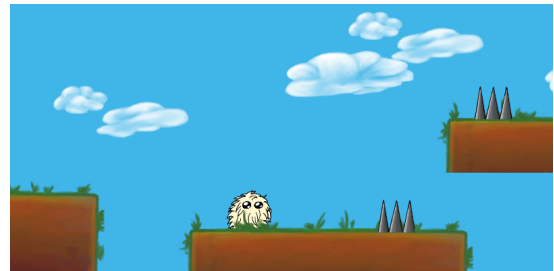

Figure 4. *Jump!* gameplay.

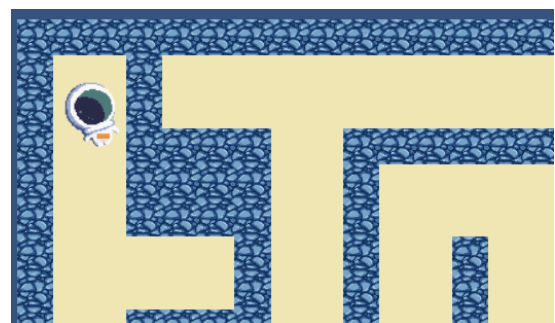**Figure 3.** *Room of the Ghosts* gameplay.

Figure 5. *Labyrinth* gameplay.

3.6 Hardware

Experimental setup consists of (see Figure 6):

- PC (processor: Intel Core i5-8600K, graphic card: MSI GeForce GTX 1070, 16 GB RAM) running under the 64-bit Windows 10 1909 Education,
- Full HD 23" LCD screen,
- External web camera Creative Live! Cam Sync HD 720p,
- Gamepad Sony PlayStation DualShock 4,
- Pioneer SE-MJ503 headphones,
- BITalino (r)evolution kit[4] with 3-leads ECG and 2-leads EDA sensors,
- Body-coloured band to hold the EDA electrodes.

Figure 6. Research setup: 23" LCD screen, headphones, external web camera, gamepad, band for EDA electrodes and Bitalino with 3-leads ECG and 2-leads EDA. Keyboard and mice were used only by the researcher to start the protocol.

3.7 Software

The procedure was running under the Python 3.8, written with PsychoPy 3.2.4 library [15]. Python code controlled the execution of the whole protocol, i.e. stimuli presentation, screencast recording (using OBS Studio[5] software), photos taking, games' start and end management.

Physiological signals were gathered using BITalino (r)evolution kit, as it is the most promising of cheap mobile hardware platforms (for comparison see [9]). Electrocardiogram recording was implemented using the classical 3 leads montage with electrodes placed below the collarbones (V– and reference) and below the last rib on the left side of the body (V+). EDA signal was gathered by 2 leads placed on the forehead (placement as good as classical palmar location [17], with no side effects related to gamepad held by the subjects). EDA electrodes were hidden under the body-coloured band, not to interfere with the facial emotion recognition process conducted by the API in the later experimental stages, and to provide tight and stable contact of the EDA electrode and skin. Both signals were probed with 1000 Hz sampling rate.

3.8 Procedure

Two research stands were prepared in the same room, located next to the opposite walls, so that the subjects sat back to back. Each participant was seated in front of a monitor and provided with the consent and short information about the whole experiment. Throughout the whole procedure, the investigator remained at his desk, rear-facing to the subjects, to reduce the Hawthrone effect.

Subjects filled out the NEO-FFI inventory and were connected to measuring devices, with the headphones set up, and a gamepad given to hand. Computer protocol consisted of five phases:

1. Baseline signals recording (1 minute),
2. Instructions and training (approx. 5 minutes),
3. First part of stimuli presentation and rating (17.5 minutes): each presentation lasted 6 seconds[6], followed by 6 seconds for affective rating and 6 second ISI[7]. Participants were instructed to navigate the procedure via gamepad.
4. Games session (up to 15 minutes in total): each game had a 5 min time limit, after which it turned itself off. After the completion of one game, another automatically switched on.
5. Second part of stimuli presentation and rating (17.5 minutes).

After the computer protocol, subjects filled out three GEQ questionnaires (one for each game) and gaming experience questionnaire.

ECG and GSR signal, as well as gamepad accelerometer and gyroscope readings, were collected continuously during the whole experiment. Facial photos were taken every 250 milliseconds. A screencast was recorded during the game session, in case for the need to fill in the missing information in game logs after the experiment. The whole protocol lasted up to 75 minutes.

4 THE BIRAFFE2 DATASET

BIRAFFE2 dataset is available to download at Zenodo[8].

It consists of a metadata file and 7 archives of the following data collections:

BIRAFFE2-metadata.csv contains a summary of each participant: age, sex, personality profile, GEQ results and information about subsets available for given person (whether there is a `BioSigs`, `Screencast`, ... file available for the person),

BIRAFFE2-biosigs.zip contains biosignals (ECG and GSR),

BIRAFFE2-gamepad.zip contains accelerometer and gyroscope recordings,

BIRAFFE2-games.zip contains logs from the games,

BIRAFFE2-photo.zip contains face emotions description calculated by MS Face API,

BIRAFFE2-photo-full.zip contains all information available in `BIRAFFE2-photo.zip` but also other face-related values recognized by MS Face API, e.g. recognized age, whether the person wears glasses, what is the color of the hair,

BIRAFFE2-procedure.zip contains a log of all the stimuli presented to a given user (timestamp of the stimuli presentation, condition, widget, stimuli ID...),

BIRAFFE2-screencast.zip contains screencasts from games.

All files have Unix timestamps which can be used for synchronization between different subsets. Detailed low-level specification of all values is provided in Sect. 4.1-4.8.

BIRAFFE2 dataset consists of data gathered from 102 out of 103 participants. Unfortunately, during the protocol for subject 723 there were problems with the hard drive and all data was lost, except the paper-and-pen NEO-FFI and GEQ questionnaires. Also, some smaller issues occurred for a subset of participants, e.g. game crashed, Bluetooth signal was lost, electrode contact was poor. We have published also the incomplete records (see Tab. 1), as in many analysis only selected of the subsets will be used and it will not be the problem. Missing values in all files are represented by `NaN`.

Table 1. Size of subsets collected (out of 103 participants)

GEQ & NEO-FFI	103
BioSigs	102
Gamepad	102
Game 1 logs	102
Game 2 logs	101
Game 3 logs	87
Photos	102
Procedure	102
Screencast	92

4.1 `BIRAFFE2-metadata.csv`

Each line of this file represents one subject and consists of the following values:

```
ID;AGE;SEX;NEO-FFI;GEQ;BIOSIGS;GAMEPAD;GAME-1;GAME
-2;GAME-3;PHOTOS;PROCEDURE;SCREENCAST;OPENNESS;
CONSCIENTIOUSNESS;NEUROTICISM;AGREEABLENESS;
EXTRAVERSION;GAME-EXO-PAST-YEAR;GAME-EXP-MOST-
INTENSE;GEQ-1-COMPETENCE-2013;GEQ-1-IMMERSION-2013;
GEQ-1-FLOW-2013;GEQ-1-TENSION-2013;GEQ-1-CHALLENGE
-2013;GEQ-1-NEGATIVE-AFFECT-2013;GEQ-1-POSITIVE-
AFFECT-2013;GEQ-1-POSITIVE-AFFECT-2018;GEQ-1-
NEGATIVITY-2018;GEQ-1-COMPETENCE-2018;GEQ-1-FLOW
-2018;GEQ-1-IMMERSION-2018;GEQ-2-COMPETENCE-2013;
GEQ-2-IMMERSION-2013;GEQ-2-FLOW-2013;GEQ-2-TENSION
-2013;GEQ-2-CHALLENGE-2013;GEQ-2-NEGATIVE-AFFECT
-2013;GEQ-2-POSITIVE-AFFECT-2013;GEQ-2-POSITIVE-
AFFECT-2018;GEQ-2-NEGATIVITY-2018;GEQ-2-COMPETENCE
-2018;GEQ-2-FLOW-2018;GEQ-2-IMMERSION-2018;GEQ-3-
COMPETENCE-2013;GEQ-3-IMMERSION-2013;GEQ-3-FLOW
-2013;GEQ-3-TENSION-2013;GEQ-3-CHALLENGE-2013;GEQ
-3-NEGATIVE-AFFECT-2013;GEQ-3-POSITIVE-AFFECT-2013;
GEQ-3-POSITIVE-AFFECT-2018;GEQ-3-NEGATIVITY-2018;
GEQ-3-COMPETENCE-2018;GEQ-3-FLOW-2018;GEQ-3-
IMMERSION-2018
```

where[9]:

- **ID** – a randomly assigned subject ID from range {100,999}. It is used to identify all subject-related files as filenames. Filenames are created according to the format `SUBxxx-yyyy`, where `xxx` is the ID, and `yyyy` is the data type identifier (e.g. `BioSigs`, `Gamepad`),
- **NEO-FFI;GEQ;BIOSIGS;GAMEPAD;GAME-1;GAME-2; GAME-3;PHOTOS;PROCEDURE;SCREENCAST** – information about subsets available for given person, i.e. whether there is a `BioSigs` file, `Gamepad` file, etc. available for the person (`Y` or `NaN`). `GAME-X` columns inform about availability of game

---

[9] Note that not all values are described in details, as some of them are obvious.

logs for level `X`. The NEO-FFI and GEQ columns only indicate whether there are questionnaire results in the following columns,

- **OPENNESS;CONSCIENTIOUSNESS;EXTRAVERSION; AGREEABLENESS;NEUROTICISM** – five personality traits calculated from raw NEO-FFI results; values represent ten scores, i.e. the possible values are in {1,2,3,…,10} set and represent standard normal distribution with $M = 5.5$ and $SD = 2$. For further analyses they can be transformed to low (1-3), medium (4-6) and high (7-10) trait levels,
- **GAME-EXO-PAST-YEAR;GAME-EXP-MOST-INTENSE** – answers from our own simple questionnaire for gaming experience measurement described in Sect. 3.2. The possible values are {A,B,C,D,E},
- **GEQ-X-Y-Z** – are calculated components values from GEQ questionnaires. X represents the game level ({1,2,3}), Y – the component name, Z – version of the GEQ (2013 is the original work [6], while 2018 is the revised version [8]). The values are ranging from 0 (not at all) to 4 (extremely) for each factor.

4.2 `BIRAFFE2-biosigs.zip`

Each `SUBxxx-BioSigs.csv` file represents one subject and consists of one line per each *sensor recording*. Values were recorded with 1 kHz frequency. The fields contained in each line are:

```
TIMESTAMP;ECG;GSR
```

where:

- **ECG** – signal (units: mV) gathered by BITalino, after units transformation recommended by BITalino Sensor Datasheet[10], low-pass filtering in 35 Hz and baseline removal performed using Python library for biosignal processing[11].
- **GSR** – signal (units: µS) gathered by BITalino, after units transformation recommended by BITalino Sensor Datasheet[12] and low-pass filtering (in range between 0.5 to 50 Hz depending on the noise level in the individual file) performed using Python library for biosignal processing[13].

The signals were recorded using the Python library for BITalino[14]. Due to the instability of the bluetooth connection, and in the absence of handling such a situation in the library, the timestamps after series[15] of several connection errors cannot be considered as fully reliable. This will be further investigated by our team.

4.3 `BIRAFFE2-gamepad.zip`

Each `SUBxxx-gamepad.csv` file represents one subject and consists of one line per each *gamepad recording*. The values were recorded as quickly as they were transmitted through the USB interface, with an average frequency of 250 Hz. The fields contained in each line are:

---

[10] See:https://bitalino.com/datasheets/ECG_Sensor_Datasheet.pdf.
[11] See:https://github.com/paulvangentcom/heartrate_analysis_python.
[12] See:https://bitalino.com/datasheets/EDA_Sensor_Datasheet.pdf.
[13] See:https://github.com/paulvangentcom/heartrate_analysis_python.
[14] See: http://bitalino.com/pyAPI/.
[15] Errors are considered to occur in the series if the interval between consecutive error entries in the **Procedure** log (see Sect. 4.7) was less than 0.05 s.

```
TIMESTAMP;GYR-X;GYR-Y;GYR-Z;ACC-X;ACC-Y;ACC-Z
```

where `GYR-` are gyroscope readings, while `ACC-` are accelerometer values. Note that the gyroscope values represent the *position* of the gamepad, not the angular rate[16]:

- **GYR-X** – right side of the gamepad upward,
- **GYR-Y** – buttons and joysticks panel upward,
- **GYR-Z** – audio port upward / light bar downward,
- **ACC-X** – yaw counter-clockwise,
- **ACC-Y** – pitch upward,
- **ACC-Z** – roll left side of gamepad down.

4.4 `BIRAFFE2-games.zip`

Five JSON log files are created for each subject. Note that the file was created when the level has started. If the game has crashed in a given level (which sometimes happened), the subsequent levels were not started and the subject was returned to the stimulus session. This means that in some cases files for may be missing, e.g. files for level 3 if the game crashed at the second level.

SUBxxx-Level01_Log.json contains a log from the *Room of the Ghosts* composed of information collected at multiple points in time about the current position and status of the user. It is a repeated pattern of the following structure:

```
1  [
2    {
3      "x": Float,  # current position on the map
4      "y": Float,
5      "deathCount": Int,
6      "shootsCounter": Int,  # both may be used to
7      "hitCounter": Int,     # calculate the accuracy
8                             # of the player
9      "money": Int,  # current score (death: -5)
10     "collectedMoney": Int,   # how many pickups
11     "collectedHealth": Int,  # were collected
12     "health": Int,
13     "timestamp": Int,  # current Unix timestamp
14     "idOfSound": String,  # ID of IADS sound played
15                           # in the background
16     "timestampOfSound": String,  # time when sound
17                                  # was started
18     "xMin": Float,  # visible area of the map
19     "yMin": Float,
20     "xMax": Float,
21     "yMax": Float
22   },
23   ...
24 ]
```

SUBxxx-Level01_MapLog.json contains information about the dynamic environment in the *Room of the Ghosts*. It is composed of series of three subsequent lists – each describing current position of existing ghosts, money bags and health pickups:

```
1  [
2    {
3      "Ghosts": [
4        {
5          "x": Float,  # current position
6          "y": Float,
7          "timestamp": Int  # Unix timestamp when the
8                            # position was recorded
```

---

16 As in the DS4Windows library, which code was used as the base for our Python data acquisition code. See: https://github.com/Jays2Kings/DS4Windows.

```
9        },
10       ... # each ghost has a separate entry
11     },
12     {
13       "Money bags": [
14         ... # the list is the same as for ghosts
15       ]
16     },
17     {
18       "Health pickups": [
19         ... # the list is the same as for ghosts
20       ]
21     },
22     ...
23 ]
```

SUBxxx-Level02_Log.json is a log from the *Jump!*, as the log from the first level, it is composed of a repeated pattern collected at multiple points in time:

```
1  [
2    {
3      "x": Float,  # current position
4      "y": Float,
5      "deathCount": Int,
6      "distortionLevel": Float,  # current parameters
7      "pitchLevel": Float,       # of the sound
8      "timestamp": Int,
9      "idOfSound": String,  # as in the first level
10     "timestampOfSound": String,
11     "xMin": Float,  # visible area of the map
12     "yMin": Float,
13     "xMax": Float,
14     "yMax": Float
15   },
16   ...
17 ]
```

SUBxxx-Level02_BlockEvents.json contains the information about the dynamic blocks of the *Jump!* game world. The design is analogous to SUBxxx-Level01_MapLog.json:

```
1  [
2    {
3      "Invisible blocks": [
4        {
5          "x": Float,
6          "y": Float,
7          "timestamp": Int
8        },
9        ... # each block has a separate entry
10     },
11     {
12       "Falling blocks": [
13         ... # the list is the same as above
14       ]
15     },
16     ...
17 ]
```

SUBxxx-Level03_Log.json is a log from the *Labyrinth*, as the logs from the previous levels, it consists of often sampled structure:

```
1  [
2    {
3      "x": Float,  # current position
4      "y": Float,
5      "deadEnds": Int,  # counts how many times
6                        # the player was off the
7                        # correct path
8      "wasOnCorrectPath": Bool,
```

```
9      "timestamp": Int,
10     "idOfSound": String,
11     "timestampOfSound": String,
12     "xMin": Float,
13     "yMin": Float,
14     "xMax": Float,
15     "yMax": Float
16   },
17   ...
18 ]
```

There are also three files containing static map of each level (same for each subject): **Level01_StaticMap.json**, **Level02_StaticMap.json**, **Level03_StaticMap.json**. They are located in the root directory of the game subset. Each of these files consists of a list of position of all (squared) blocks building the maps:

```
1 [
2   { "x": Float, "y": Float, "z": Float },  # 1st
3   { "x": Float, "y": Float, "z": Float },  # 2nd
4   { "x": Float, "y": Float, "z": Float },  # 3rd
5   ...
6 ]
```

4.5 **BIRAFFE2-photo.zip**

Each `SUBxxx-Face.csv` file represents one subject and consists of one line per each *photo taken*. Raw photos are not available due to privacy reasons. File consists of values calculated by MS Face API with *recognition_02* model. Photos were taken with 4 Hz frequency during games and during stimuli presentation (every 15 frames at 60 fps). Photos were not taken while the subject was responding on the widget. When no face was recognized or two faces were found (the second was the experimenter face) `NaN` value was used.

```
GAME-TIMESTAMP;FRAME-NUMBER;IADS-ID;IAPS-ID;ANGER;
CONTEMPT;DISGUST;FEAR;HAPPINESS;NEUTRAL;SADNESS;
SURPRISE
```

where:

- **GAME-TIMESTAMP** – Unix timestamp available only during the game (`NaN` value during the stimuli presentation),
- **FRAME-NUMBER** – Index of the photo within the context of the stimuli presentation (`NaN` value during the games), measured in frames since the beginning of the stimuli presentation: -1 for pre-stimulation photo, 0 for photo in the moment when stimuli appears, 15 for the next photo ($1/4$ s later), up to 345 (frame 360 = 6 s = time when stimuli disappears),
- **IADS-ID;IAPS-ID** – IADS/IAPS IDs of stimuli (see Sect. 4.7),
- **ANGER;CONTEMPT;DISGUST;FEAR;HAPPINESS;NEUTRAL;SADNESS;SURPRISE** – probability distribution of eight emotions calculated by MS Face API (all values sum up 1). It is important to note that this distribution is highly skewed to the NEUTRAL emotion, having values close to 1 in that emotion and values close to zero in the rest of them.

4.6 **BIRAFFE2-photo-full.zip**

`SUBxxx-Face.csv` files are analogous to the `BIRAFFE-photo.zip`, but with the full output from MS Face API. Besides the values described in Sect. 4.5, they also have the following face-related values[17]:

```
FACEATTRIBUTES-ACCESSORIES;FACEATTRIBUTES-AGE;
FACEATTRIBUTES-BLUR-BLURLEVEL;FACEATTRIBUTES-BLUR-
VALUE;FACEATTRIBUTES-EXPOSURE-EXPOSURELEVEL;
FACEATTRIBUTES-EXPOSURE-VALUE;FACEATTRIBUTES-
FACIALHAIR-BEARD;FACEATTRIBUTES-FACIALHAIR-
MOUSTACHE;FACEATTRIBUTES-FACIALHAIR-SIDEBURNS;
FACEATTRIBUTES-GENDER;FACEATTRIBUTES-GLASSES;
FACEATTRIBUTES-HAIR-BALD;FACEATTRIBUTES-HAIR-
HAIRCOLOR-BLACK;FACEATTRIBUTES-HAIR-HAIRCOLOR-BLOND
;FACEATTRIBUTES-HAIR-HAIRCOLOR-BROWN;FACEATTRIBUTES
-HAIR-HAIRCOLOR-GRAY;FACEATTRIBUTES-HAIR-HAIRCOLOR-
OTHER;FACEATTRIBUTES-HAIR-HAIRCOLOR-RED;
FACEATTRIBUTES-HAIR-INVISIBLE;FACEATTRIBUTES-
HEADPOSE-PITCH;FACEATTRIBUTES-HEADPOSE-ROLL;
FACEATTRIBUTES-HEADPOSE-YAW;FACEATTRIBUTES-MAKEUP-
EYEMAKEUP;FACEATTRIBUTES-MAKEUP-LIPMAKEUP;
FACEATTRIBUTES-NOISE-NOISELEVEL;FACEATTRIBUTES-
NOISE-VALUE;FACEATTRIBUTES-SMILE;FACEID;
FACELANDMARKS-EYEBROWLEFTINNER-X;FACELANDMARKS-
EYEBROWLEFTINNER-Y;FACELANDMARKS-EYEBROWLEFTOUTER-X
;FACELANDMARKS-EYEBROWLEFTOUTER-Y;FACELANDMARKS-
EYEBROWRIGHTINNER-X;FACELANDMARKS-EYEBROWRIGHTINNER
-Y;FACELANDMARKS-EYEBROWRIGHTOUTER-X;FACELANDMARKS-
EYEBROWRIGHTOUTER-Y;FACELANDMARKS-EYELEFTBOTTOM-X;
FACELANDMARKS-EYELEFTBOTTOM-Y;FACELANDMARKS-
EYELEFTINNER-X;FACELANDMARKS-EYELEFTINNER-Y;
FACELANDMARKS-EYELEFTOUTER-X;FACELANDMARKS-
EYELEFTOUTER-Y;FACELANDMARKS-EYELEFTTOP-X;
FACELANDMARKS-EYELEFTTOP-Y;FACELANDMARKS-
EYERIGHTBOTTOM-X;FACELANDMARKS-EYERIGHTBOTTOM-Y;
FACELANDMARKS-EYERIGHTINNER-X;FACELANDMARKS-
EYERIGHTINNER-Y;FACELANDMARKS-EYERIGHTOUTER-X;
FACELANDMARKS-EYERIGHTOUTER-Y;FACELANDMARKS-
EYERIGHTTOP-X;FACELANDMARKS-EYERIGHTTOP-Y;
FACELANDMARKS-MOUTHLEFT-X;FACELANDMARKS-MOUTHLEFT-Y
;FACELANDMARKS-MOUTHRIGHT-X;FACELANDMARKS-
MOUTHRIGHT-Y;FACELANDMARKS-NOSELEFTALAROUTTIP-X;
FACELANDMARKS-NOSELEFTALAROUTTIP-Y;FACELANDMARKS-
NOSELEFTALARTOP-X;FACELANDMARKS-NOSELEFTALARTOP-Y;
FACELANDMARKS-NOSERIGHTALAROUTTIP-X;FACELANDMARKS-
NOSERIGHTALAROUTTIP-Y;FACELANDMARKS-
NOSERIGHTALARTOP-X;FACELANDMARKS-NOSERIGHTALARTOP-Y
;FACELANDMARKS-NOSEROOTLEFT-X;FACELANDMARKS-
NOSEROOTLEFT-Y;FACELANDMARKS-NOSEROOTRIGHT-X;
FACELANDMARKS-NOSEROOTRIGHT-Y;FACELANDMARKS-NOSETIP
-X;FACELANDMARKS-NOSETIP-Y;FACELANDMARKS-PUPILLEFT-
X;FACELANDMARKS-PUPILLEFT-Y;FACELANDMARKS-
PUPILRIGHT-X;FACELANDMARKS-PUPILRIGHT-Y;
FACELANDMARKS-UNDERLIPBOTTOM-X;FACELANDMARKS-
UNDERLIPBOTTOM-Y;FACELANDMARKS-UNDERLIPTOP-X;
FACELANDMARKS-UNDERLIPTOP-Y;FACELANDMARKS-
UPPERLIPBOTTOM-X;FACELANDMARKS-UPPERLIPBOTTOM-Y;
FACELANDMARKS-UPPERLIPTOP-X;FACELANDMARKS-
UPPERLIPTOP-Y;FACERECTANGLE-HEIGHT;FACERECTANGLE-
LEFT;FACERECTANGLE-TOP;FACERECTANGLE-WIDTH
```

4.7 **BIRAFFE2-procedure.zip**

Each `SUBxxx-Procedure.csv` file represents one subject and consists of one line per each *stimuli presentation*. The fields contained in each line are:

```
TIMESTAMP;ID;COND;IADS-ID;IAPS-ID;ANS-VALENCE;ANS-
AROUSAL;ANS-TIME;EVENT
```

where:

- **TIMESTAMP** – Unix timestamp when the stimuli appeared on the screen,
- **ID** – subject ID,

---

[17] See also MS Face API documentation at `https://westus.dev.cognitive.microsoft.com/docs/` `services/563879b61984550e40cbbe8d/operations/` `563879b61984550f30395236`.

- **COND** – one of nine conditions as specified in Sect. 3.3 (`P+`, `P0`, `P-`, `S+`, `S0`, `S-`, `P+S+`, `P0S0`, `P-S-`),
- **IADS-ID;IAPS-ID** – IADS/IAPS IDs of stimuli. Both IADS and IAPS datasets provide Valence/Arousal scores for each stimuli that can be used for further analyses (these values describe emotions that were evoked by the stimuli). Contact with the CSEA at University of Florida to obtain your own copy of the datasets for research[18],
- **ANS-VALENCE;ANS-AROUSAL** – values in $[1;9]$ ranges indicating the point selected by the subject in the *Valence-arousal faces* widget (see Sect. 3.4),
- **ANS-TIME** – response time (0 is a moment when widget appeared on the screen); NaN indicates that the subject has not made any choice but left the default option,
- **EVENT** – information about going through the next procedure checkpoint (e.g. tutorial start, game session end) and about BITalino errors (see Sect. 4.2).

4.8 `BIRAFFE2-screencast.zip`

Each `SUBxxx-Screencast.mkv` file contains a *screen recording* (1920x1080 resolution, 60 fps, h.264 codec) of game session for one subject. The entire subset weights 12 GB and, due to its large size, it is available on request.

5 SUMMARY AND FUTURE WORK

This paper describes the BIRAFFE2 data set collected in 2020 during the affective computing experiment, aiming to provide a framework for the enhanced models of emotion classification an recognition. The enhancement, founded by the variety of collected data types and behavioural features, is grounded in the belief, that the up-to-date emotion recognition models should be personalized by design. The construction of such models, as well as the experimental design used for data collection, should take the individual personality differences into account and provide a solution for its operationalization. We believe, that the data set described in this paper presents an important contribution that supports the development and replication of experiments in affective computing and Human-AI interaction.

Based on the data acquired in the reported experiment we now plan a series of analyses aimed at the development of personalized models of emotion. We are planning to analyze the data from the first phase of the experiment to develop certain calibration and personalization methods to fine tune the interpretation of emotion of in several tasks. While we are currently experimenting with video games, we are also considering the incorporation of the developed models into decision support and recommendation systems.

REFERENCES

[1] *Handbook of Emotions*, eds., Lisa Feldman Barrett, Michael Lewis, and Jeannette M. Haviland-Jones, The Guilford Press, New York, NY, 4th edn., 2016.
[2] M. M. Bradley and P. 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] *The Oxford Handbook of Affective Computing*, eds., Rafael A. Calvo, Sidney K. D'Mello, Jonathan Gratch, and Arvid Kappas, Oxford Library of Psychology, Oxford University Press, Oxford, 2015.
[4] P.T. Costa and R.R. McCrae, *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI). Professional manual*, Psychological Assessment Resources, Odessa, FL, 1992.
[5] Kristina Höök, 'Affective loop experiences - what are they?', in *Persuasive Technology, Third International Conference, PERSUASIVE 2008, Oulu, Finland, June 4-6, 2008. Proceedings*, eds., Harri Oinas-Kukkonen, Per F. V. Hasle, Marja Harjumaa, Katarina Segerståhl, and Peter Øhrstrøm, volume 5033 of *Lecture Notes in Computer Science*, pp. 1–12. Springer, (2008).
[6] W.A. IJsselsteijn, Y.A.W. de Kort, and K. Poels, *The Game Experience Questionnaire*, Technische Universiteit Eindhoven, 2013.
[7] William James, 'What is an emotion?', *Mind*, **9**(34), 188–205, (1884).
[8] 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).
[9] 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 2018, Zakopane, Poland, June 3-7, 2018, Proceedings, Part II*, pp. 168–178, (2018).
[10] 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*, CEUR Workshop Proceedings. CEUR-WS.org, (2020).
[11] P. J. Lang, M. 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).
[12] Effie L.-C. Law, Florian Brühlmann, and Elisa D. Mekler, 'Systematic review and validation of the game experience questionnaire (geq) - implications for citation and reporting practice', in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, p. 257–270, New York, NY, USA, (2018). Association for Computing Machinery.
[13] Grzegorz J. Nalepa, Krzysztof Kutt, and Szymon Bobek, 'Mobile platform for affective context-aware systems', *Future Generation Computer Systems*, **92**, 490–503, (mar 2019).
[14] 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).
[15] 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).
[16] Rosalind W. Picard, *Affective Computing*, MIT Press, Cambridge, MA, 1997.
[17] Marieke van Dooren, J. J. G. de Vries, and Joris H. Janssen, 'Emotional sweating across the body: Comparing 16 different skin conductance measurement locations', *Physiology & Behavior*, **106**(2), 298 – 304, (2012).
[18] B. Zawadzki, J. Strelau, P. Szczepaniak, and M. Śliwińska, *Inwentarz osobowości NEO-FFI Costy i McCrae. Polska adaptacja*, Pracowania Testów Psychologicznych PTP, Warszawa, 1998.
[19] Laura Żuchowska, *Game Design with Unity for Affective Games*, BSc thesis, AGH University of Science and Technology, 2020. Supervisor: G.J. Nalepa.

---

[18] See: `https://csea.phhp.ufl.edu/media.html`.