People in the Context – an Analysis of Game-based Experimental Protocol

Krzysztof Kutt[1*] , **Laura Żuchowska**[2] , **Szymon Bobek**[1,2] and **Grzegorz J. Nalepa**[1,2]
[1]Jagiellonian Human-Centered Artificial Intelligence Laboratory (JAHCAI) and Institute of Applied Computer Science, Jagiellonian University, Kraków, Poland
[2]Department of Applied Computer Science, AGH University of Science and Technology, Kraków, Poland
krzysztof.kutt@uj.edu.pl, szymon.bobek@uj.edu.pl, gjn@gjn.re

Abstract

The paper provides insights into two main threads of analysis of the BIRAFFE2 dataset concerning the associations between personality and physiological signals and concerning the game logs' generation and processing. Alongside the presentation of results, we propose the generation of event-marked maps as an important step in the exploratory analysis of game data. The paper concludes with a set of guidelines for using games as a context-rich experimental environment.

1 Introduction and Motivation

The development of a good personalised intelligent assistant that behaves in a natural way requires the development of proper toolbox as a base [Nalepa *et al.*, 2019]. In order to be user-friendly, an assistant should not only perform its task, but also respond to the user's changing emotions. This is due to our natural tendency to anthropomorphize interfaces – the user will assume that the assistant will react appropriately, e.g., understand that the nervousness is due to a mistake committed. Such affective information can be extracted from the range of physiological signals, particularly obtained through low-cost wearable devices that will make this technology available to everyone. Finally, it is important to note that emotions do not happen in a void—they are always dependent on the context a person is in [Prinz, 2006]—so it is also important to collect information about the user's current situation (e.g., activity, weather conditions, time of day).

An important step in establishing the above-outlined framework for personalized assistants is the collection of the right data. This, in turn, strictly depends on the development of appropriate research environments and experimental protocols. Such issues are addressed in the BIRAFFE (*Bio-Reactions and Faces for Emotion-based Personalization*) series of experiments [Kutt *et al.*, 2021]. Their main objective is to develop methods for emotion recognition using a range of contextual information and physiological signals such as cardiac activity (ECG), electrodermal response (EDA), hand movements (accelerometer) or changes in facial expression. In order to ensure that the research is highly ecological in measurement and easily extendable to wider research groups, wearable and portable, affordable-for-all devices are used.

A key aspect of the BIRAFFE experiments is the use of games as the experimental environment. They were chosen as a trade-off between a stimulus-rich complex near-natural environment and the need to control and record as much context as possible to provide the most detailed post-experimental analyses. The latest version of the experiment, BIRAFFE2 [Kutt *et al.*, 2020][1], used a game consisting of three independent levels. The aim of the first was to evoke positive emotions. The second was intended to induce irritation and frustration, e.g., through impaired control. Finally, the third level was a neutral maze. A detailed description of the games is presented in [Żuchowska *et al.*, 2020].

This paper provides insights into the core analyses of the BIRAFFE2 dataset on contextual information processing in affective games. The first thread, presented in Sect. 2, focuses on the analysis of the relationship between physiological signals and the so-called "Big Five" personality traits. The existence of such relationships in the data will allow further work to create emotion prediction models that will be moderated and personalised through the identification of personality profiles. The second topic, described in Sect. 3, addresses the topic of accurate game logging and the possibility of reconstructing an entire game from such stored logs. The whole article concludes with a set of lessons-learned regarding the implementation of games as an experimental environment in Sect. 4.

2 Physiological Signals and Personality

Before undertaking the analyses, three features were calculated for ECG signal using HeartPy library [van Gent *et al.*, 2019]: heart rate (number of heart beats per minute), mean of successive differences between R-R intervals (MoSD) and breathing rate. Also, to group the valence and arousal scores into discrete variable, 16 clusters were introduced as presented on Fig. 1.

In order to find correlations and dependencies between physiological data (the ECG signal was chosen as an illustration) and personality traits (each on $[1, 10]$ scale), several

---

[*] Corresponding Author

[1] The entire dataset from the BIRAFFE2 experiment is available under CC licence on the Zenodo platform, DOI:10.5281/zenodo.3865859.

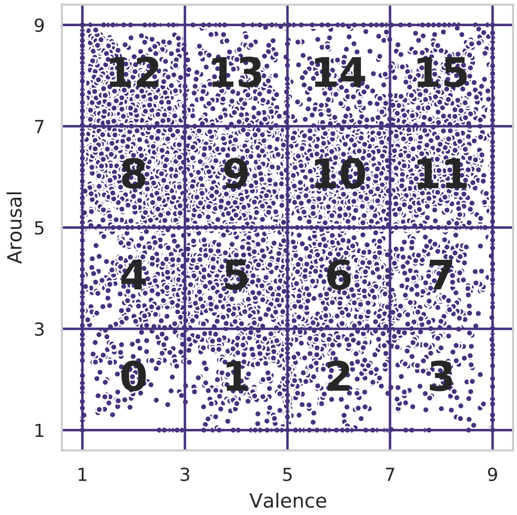


Figure 1: Valence and arousal scores grouped into clusters.

Personality Trait	Mean	SD	Median
Conscientiousness	5.68	2.61	6
Openness	5.48	2.22	6
Agreeableness	6.25	2.39	6
Extraversion	5.84	2.29	7
Neuroticism	5.38	2.45	5

Table 1: Descriptive statistics for personality traits.

approaches to statistical analysis were made. Firstly, basic descriptive statistics were calculated to find outliers and possible extremas. As can be seen in Tab. 1-2, the data was distributed proportionally in terms of mean, median and standard deviation, which indicates a promising start for further analysis.

The second analysis was aimed at investigation of correlations between features. Although the results did not show any strong dependencies between them (see Fig. 2), they indicated the existence of potentially interesting relationships worthy of further analysis and further research. Namely, in terms of the associations between personality and widget responses, valence and arousal are related to distinct traits. For arousal, the highest values are for openness and conscientiousness. On the other hand, valence's most significant factors are agreeableness and extraversion. When considering the correlations between physiological reactions and widget, among heart rate, MoSD, and respiratory rate, the highest values were noted for the first of these for both valence and arousal. The outcome of personality trait to heart rate was presented as maximal for both conscientiousness and extraversion. Considering the MoSD, highest value—and the highest inter-correlation in general, i.e., the correlation between different data sources—was for extraversion (0.23) and conscientiousness (-0.19). Finally, values of correlation for breathing rate played in favor of extraversion.

ECG characteristic	Mean	SD
Heart rate [BPM]	80.92	16.21
MoSD [ms]	34.41	37.89
Breathing rate [Hz]	0.10	0.12

Table 2: Descriptive statistics for ECG characteristics.

Independent var.	df	MS	F	p
Conscientiousness	1	3.85	0.60	0.44
Openness	1	5.24	0.82	0.37
Agreeableness	1	68.57	10.70	**0.001**
Extraversion	1	6.12	0.95	0.33
Neuroticism	1	0.02	0.004	0.95
Heart rate	1	59.97	9.35	**0.002**
MoSD	1	0.87	0.14	0.71
Breathing rate	1	2.91	0.45	0.50
Residual Error	10881	6.41		

Table 3: ANOVA model for *valence* with five personality traits and three ECG-related characteristics as independent variables.

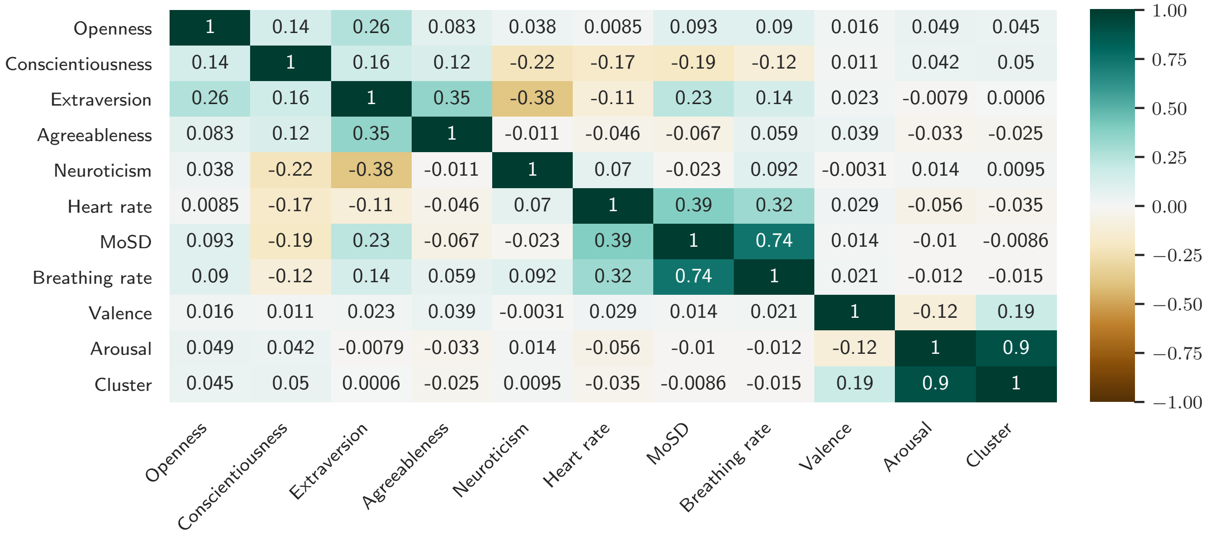
The last statistical analysis performed was two ANOVAs for valence and arousal (see Tab. 3-4), which indicated several strong associations. What seems most interesting is the strong relationship between heart rate and valence, which is somehow in opposition to most approaches in which heart rate is used to predict arousal, while other signals such as EDA are mostly used for valence [Dzedzickis *et al.*, 2020].

3 Game Logs and Questionnaires

As noted in the introduction, one motivation for using games as an experimental environment is the ability to frequently sample and log the entire player context. Properly prepared logs should allow the reconstruction of both the level map (the same for each subject) and the course of the entire game for each player. Indeed, this is possible for the games studied. As part of the log analyses, a number of maps were generated, which were verified by comparison with the games and recorded screencasts of the gameplay. These maps can also be used for aggregated analyses, e.g., by plotting all events of one type followed by an initial visual inspection. Fig. 3 shows all the death locations of the protagonist in the first level. One can notice a very high number of deaths in the central room – this is consistent with the observations made during the experiment: this is the first room where players are just getting familiar with the game interface.

Another part of the analysis was the examination of answers from Game Experience Questionnaire [IJsselsteijn *et al.*, 2013], a survey taken by each participant by the end of the experiment. The results allow to understand whether the games made an impact on emotional state of the subjects, according to themselves. The results are represented by 7-factor structure. Five of them were further analysed, as they were the most relevant to the assumed game differences:

- Challenge – I felt time pressure/I had to put a lot effort,
- Tension – I was irritated/I feel angry,
- Negative affect – I felt bad/made me bored,



	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	Heart rate	MoSD	Breathing rate	Valence	Arousal	Cluster
Openness	1	0.14	0.26	0.083	0.038	0.0085	0.093	0.09	0.016	0.049	0.045
Conscientiousness	0.14	1	0.16	0.12	-0.22	-0.17	-0.19	-0.12	0.011	0.042	0.05
Extraversion	0.26	0.16	1	0.35	-0.38	-0.11	0.23	0.14	0.023	-0.0079	0.0006
Agreeableness	0.083	0.12	0.35	1	-0.011	-0.046	-0.067	0.059	0.039	-0.033	-0.025
Neuroticism	0.038	-0.22	-0.38	-0.011	1	0.07	-0.023	0.092	-0.0031	0.014	0.0095
Heart rate	0.0085	-0.17	-0.11	-0.046	0.07	1	0.39	0.32	0.029	-0.056	-0.035
MoSD	0.093	-0.19	0.23	-0.067	-0.023	0.39	1	0.74	0.014	-0.01	-0.0086
Breathing rate	0.09	-0.12	0.14	0.059	0.092	0.32	0.74	1	0.021	-0.012	-0.015
Valence	0.016	0.011	0.023	0.039	-0.0031	0.029	0.014	0.021	1	-0.12	0.19
Arousal	0.049	0.042	-0.0079	-0.033	0.014	-0.056	-0.01	-0.012	-0.12	1	0.9
Cluster	0.045	0.05	0.0006	-0.025	0.0095	-0.035	-0.0086	-0.015	0.19	0.9	1

Figure 2: Correlation matrix for five personality traits, three ECG-related characteristics and widget responses represented as Valence, Arousal and Cluster.

Independent var.	df	MS	F	p
Conscientiousness	1	60.43	14.50	**< 0.001**
Openness	1	101.03	24.25	**< 0.001**
Agreeableness	1	44.51	10.68	**0.001**
Extraversion	1	9.61	2.31	0.13
Neuroticism	1	13.14	3.15	0.08
Heart rate	1	138.06	33.13	**< 0.001**
MoSD	1	11.27	2.71	0.10
Breathing rate	1	1.64	0.39	0.53
Residual Error	10881	4.17		

Table 4: ANOVA model for *arousal* with five personality traits and three ECG-related characteristics as independent variables.

- Positive affect – I felt good/made me happy,
- Competence – I felt competent/skillful.

The factors were compared to each other in order to dig into the feelings of players. The expectations for the first game were that subject is supposed to feel happy (high positive, low negative, low tension) and not challenged (high competence, low challenge). The second stage's purpose was contrary to the first one – high negative, tension and challenge, with low competence and positive. The huge difference is more likely to have an impact, as the contrast is hitting the player suddenly. Based on the GEQ results (see Fig. 4), one can state that everything worked as planned.

The Competence line during first gameplay was set pretty high, while leaving the tension line in the bottom, making the subject feel calm enough to let their guard down, but still be entertained by the gameplay. The second stage's extreme difficulty and pressure-building environment made the experience hard to enjoy. A very similar result can be seen in Negative/Tension comparison. About 95% of the participants agreed that the second level has left them irritated, 83,5% were not happy during and after the game. This cannot be said about the first stage, where according to the answers, only 30% of subjects felt somewhat irritated. Same outcome can be said about positive feedback for both stages – the first was keeping the emotions of participants on a very high level of happiness, while the second one changed it for a little one.

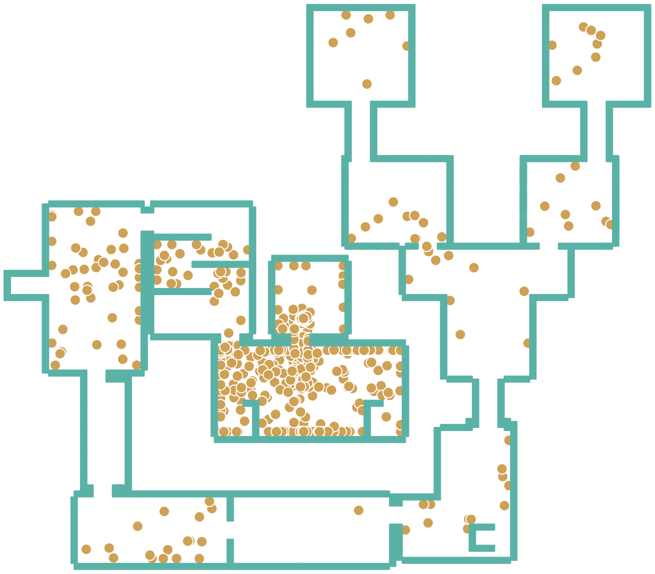


Figure 3: Map for Stage 1 recreated from game logs with all death-related events marked as dots.

4 Discussion and Lessons Learned

As a summary of the analyses presented, we propose a set of guidelines concerning the issues one should pay attention to when creating games with the intention of using them as context-rich experimental environments:

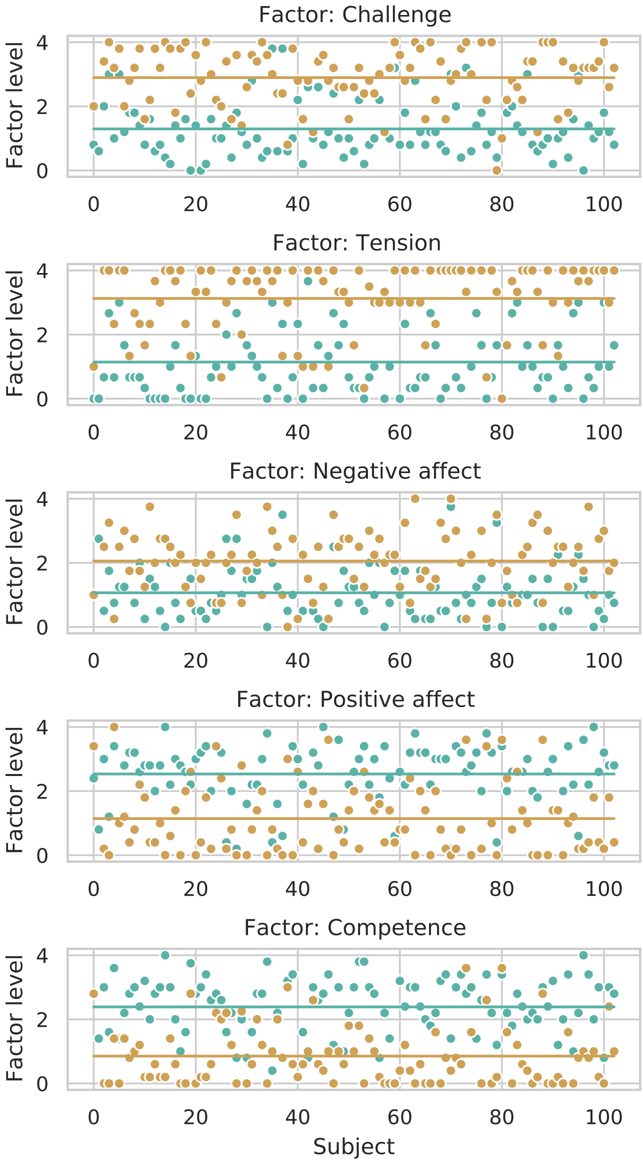


Figure 4: GEQ factors for the first and second level (green and yellow, respectively). Horizontal lines mark the average values.

1. *It is important to take into account the features of the subjects in the contextual information set.* In line with the results obtained from the BIRAFFE1 [Kutt *et al.*, 2021] and DEAP [Zhao *et al.*, 2019] datasets, the analyses summarised in Sect. 2 indicate interesting relationships between personality traits and physiological signals. Merging such several subject-related contextual information will allow a more accurate analysis leading to better modelling of a person's behaviour in the considered environment.

2. *The set of stimuli should be well balanced* so that there are neither too many (which will make analysis difficult) nor too few (the environment will not be interesting for the subject). Small levels, each focusing on selected aspects, should be preferred to one large level that combines all experimental manipulations. The levels analysed achieved their objectives well, as shown by the results of the GEQ questionnaire in Sect. 3.

3. *Logs should be collected as densely as possible, according to the specifics of the game being developed.* All features necessary to reproduce the gameplay should be recorded. In the analyses carried out, it was found that the logs were sufficiently detailed to reproduce the progress of the game. However, the data lacked information on the type of death in the second level, which would be useful to compare with the emotions felt at the time of death. This information is still reproducible, e.g., from the recorded screencasts, however it will require a fair amount of data processing.

4. *Maps with events marked on them are a useful tool for exploratory analysis of game logs.* There are a number of studies concerning the analysis of game logs (e.g., [Cheong *et al.*, 2008]), including those related to the evaluation of social science theories [Shim *et al.*, 2011]. However, to the best of our knowledge, data visualisation in the form of maps (as in Fig. 3) has not been done as part of the analyses. We believe that this is a valuable approach to quickly assess the validity of the data and to propose hypotheses that have not been considered before.

These findings will be incorporated into the preparation of the next experiment in the BIRAFFE series, planned for Autumn 2021.

Acknowledgements

The research has been supported by a grant from the Priority Research Area Digiworld under the Strategic Programme Excellence Initiative at the Jagiellonian University.

The authors are also grateful to Academic Computer Centre CYFRONET AGH and Jagiellonian University for granting access to the computing infrastructure built in the projects No. POIG.02.03.00-00-028/08 "PLATON – Science Services Platform" and No. POIG.02.03.00-00-110/13 "Deploying high-availability, critical services in Metropolitan Area Networks (MAN-HA)".

References

[Cheong *et al.*, 2008] Yun-Gyung Cheong, Arnav Jhala, Byung-Chull Bae, and Robert Michael Young. Automatically generating summary visualizations from game logs. In Christian Darken and Michael Mateas, editors, *AIIDE 2008*. The AAAI Press, 2008.

[Dzedzickis *et al.*, 2020] Andrius Dzedzickis, Arturas Kaklauskas, and Vytautas Bucinskas. Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3):592, 2020.

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

[Kutt *et al.*, 2020] Krzysztof Kutt, Dominika Drążyk, Maciej Szelążek, Szymon Bobek, and Grzegorz J. Nalepa. The BIRAFFE2 experiment. study in bio-reactions and faces for emotion-based personalization for AI systems. *CoRR*, abs/2007.15048, 2020.

[Kutt *et al.*, 2021] Krzysztof Kutt, Dominika Drążyk, Szymon Bobek, and Grzegorz J. Nalepa. Personality-based affective adaptation methods for intelligent systems. *Sensors*, 21(1):163, 2021.

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

[Prinz, 2006] Jesse J. Prinz. *Gut Reactions. A Perceptual Theory of Emotion*. Oxford University Press, Oxford, 2006.

[Shim *et al.*, 2011] Kyong Jin Shim, Nishith Pathak, Muhammad Aurangzeb Ahmad, Colin DeLong, Zoheb Borbora, Amogh Mahapatra, and Jaideep Srivastava. Analyzing human behavior from multiplayer online game logs: A knowledge discovery approach. *IEEE Intell. Syst.*, 26(1):85–89, 2011.

[van Gent *et al.*, 2019] Paul van Gent, Haneen Farah, Nicole van Nes, and Bart van Arem. Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project. *Journal of Open Research Software*, 7(1):32, 2019.

[Zhao *et al.*, 2019] Sicheng Zhao, Amir Gholaminejad, Guiguang Ding, Yue Gao, Jungong Han, and Kurt Keutzer. Personalized emotion recognition by personality-aware high-order learning of physiological signals. *ACM Trans. Multim. Comput. Commun. Appl.*, 15(1s):14:1–14:18, 2019.

[Żuchowska *et al.*, 2020] Laura Żuchowska, Krzysztof Kutt, Krzysztof Geleta, Szymon Bobek, and Grzegorz J. Nalepa. Affective games provide controlable context. proposal of an experimental framework. In Jörg Cassens, Rebekah Wegener, and Anders Kofod-Petersen, editors, *Proceedings of the Eleventh International Workshop Modelling and Reasoning in Context co-located with the 24th European Conference on Artificial Intelligence, MRC@ECAI 2020, Santiago de Compostela, Galicia, Spain, August 29, 2020*, volume 2787 of *CEUR Workshop Proceedings*, pages 45–50. CEUR-WS.org, 2020.