Emotional Mario Task at MediaEval 2021

Mathias Lux[1], Michael Riegler[2], Steven Hicks[2], Duc-Tien Dang-Nguyen[3,4], Kristine Jorgensen[3], Vajira Thambawita[2], Pål Halvorsen[2]
[1]Alpen-Adria Universität Klagenfurt, Austria
[2]SimulaMet, Norway
[3]University of Bergen, Norway
[4]Kristiania University College, Norway
mlux@itec.aau.at,michael@simula.no,steven@simula.no,ductien.dangnguyen@uib.no,kristine.jorgensen@uib.no
vajira@simula.no,paalh@simula.no

ABSTRACT

Video games are often understood as *engines of experience*, and the interaction with the game lets players consume carefully constructed experiences. While it is generally agreed upon that a good experience makes a good game, methods for measuring or observing the impact of the gameplay on the players' experience are still an open problem. In the 2021 Emotional Mario task, we ask researchers to investigate the gameplay of ten study participants on one of the most iconic classic video games: Super Mario Bros. We provide data to learn from, including heart rate, skin conductivity, videos of the players' faces synchronized to the gameplay, the gameplay itself, and player demographics including their scores and times spent in the game. Participants of the task are asked to predict gameplay events based on the biometric and facial data of the players.

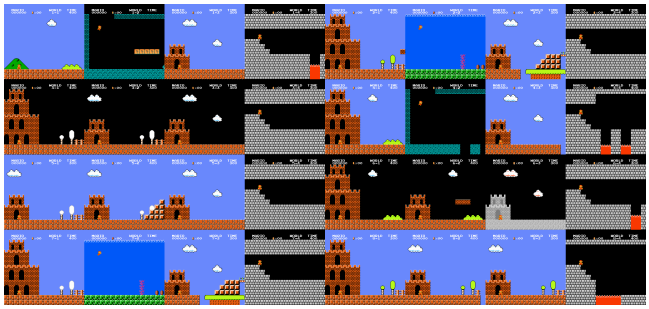

Figure 1: Collage of the first frames of all 32 *Super Mario Bros.* levels organized in eight worlds with four levels each.

1 INTRODUCTION

With the rise of deep learning, several large leaps in research have been achieved in recent years such as human-level image recognition, text classification, and even content creation. Games and deep learning also have a relatively long history together, specifically in reinforcement learning. However, video games still pose a lot of challenges. Games are understood as engines of experience [9], and as such, they need to invoke human emotions. While emotion recognition has come a far way over the last decade [7], the connection between emotions and video games is still an open and interesting research question. As games are designed to evoke emotions [9], we hypothesize that emotions in the player are reflected in the visuals of the video game. Simple examples are when players are happy after having mastered a particularly complicated challenge, when they are shocked by a jump scare scene in a horror game, or when they are excited after unlocking a new resource. Questionnaires can measure these things after playing [1], but in the Emotional Mario task, we want to interconnect emotions and gameplay based on data instead of asking the players.

For the Emotional Mario challenge, we focus on the iconic Super Mario Bros. video game and provide a multimodal dataset based on a Super Mario Bros. implementation for OpenAI Gym [2]. For a population of ten players, the dataset contains their game input, demographics, biomedical, sensory input from a medical-grade device, and videos of their faces while playing the game.

2 TASK DESCRIPTION

The goal in the Emotional Mario task is to relate data about the players, e.g., their heart rate, skin conductivity, or their facial expressions, to the gameplay and events like Mario losing a life, finishing a level, or gaining a power-up by consuming a mushroom. Emotional Mario is structured into two subtasks:

- In the first subtask, we asked participants to identify events of high significance in the gameplay by just analyzing the facial video and the biometric data. Such significant events include the end of a level, a power-up or extra life for Mario, or Mario's death.
- For the second subtask, which was optional, we asked participants to create a video summary of the best moments of the play. This can include gameplay scenes, facial video, data visualization, and whatever can help such a summary.

3 DATASET

The task provides a dataset of videos and sensor readings of people playing Super Mario Bros [8]. In total, a population of ten people was selected for data gathering, ranging from gaming veterans to novice players, with an even split between male and female participants. Each participant provided a written form of consent, allowing for their video, gameplay data, and sensor data to be shared openly for research and teaching purposes under a *Creative Commons Attribution-NonCommercial 4.0 International License*[1].

MediaEval'21, December 13-15 2021, Online

[1]http://creativecommons.org/licenses/by-nc/4.0/,accessed2020-11-04

The dataset can be accessed via https://datasets.simula.no/toadstool or https://osf.io/qrkcf/.

For each participant, a range of multimodal data was recorded and included in the dataset:

- a video file recording the participants face with a 1.3-MP webcam with 30fps and 640x480 pixels,
- the controller input performed on each frame of the game utilizing a wired USB controller from retro-bit, which is modeled after the original controller for the Nintendo Entertainment System,
- sensor data collected from an Empatica E4 wristband [4] including heart rate, temperature, skin conductivity, and accelerometer data, and
- video game action files, which are scripts to generate the video game frames.

Besides the actual data, the provided dataset includes documentation and process description as well. Additional data ranges from the original questionnaire presented to the participants and their answers, the consent form signed by the participants, the license, and a README.txt file detailing the use of the dataset. A detailed description of the dataset is given in [8].

In addition to the dataset, we provide (i) ground truth data for the events in the game for 7 out of 10 participants (the remaining three are used for the test set), and (ii) results from an automated facial expression recognition package [3] including a confidence value for the basic emotions anger, disgust, fear, happiness, sadness, and surprise, as well as a neutral expression along with a bounding box for the detected face. Examples are shown in Fig. 2.

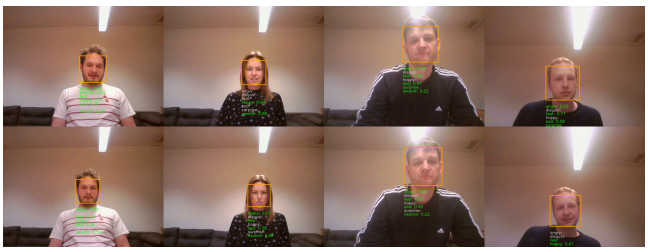

Figure 2: Sample output from the facial expression recognition algorithm with bounding box and prediction. Videos taken from the Toadstool data set [8].

4 EVALUATION

The evaluation of the task is two-fold. For the first subtask, we collect the participants' output on finding events for the missing participants. We investigated the precision, recall, and f1 score of the events. We ran four different types of evaluations. We investigated if participants found the events in a range of +/- one second of the actual events and did the same for +/- five seconds. Two more evaluations were done focusing on the time of the event, discarding the type. A random baseline was created by simulating runs with randomized events. The random baseline is biased by the knowledge of how many events are expected and chooses the number of events randomly in the range of actual number of events +/- 50%[2]. Table 1 and Table 2 give an overview of the best results of each group as well as the averaged random baseline.

Table 1: Best evaluation results per group per evaluation and the averaged random baseline for matching event timestamps in the range of +/- 5 seconds

Team	Precision	Recall	F1 score
DCU-Gamestreamer	0.0021	0.8903	0.0041
GSE-AAU	0.0242	0.0812	0.0373
Random	0.2847	0.2847	0.2947

Table 2: Best evaluation results per group per evaluation and the averaged random baseline for matching events in the range of +/- 5 seconds

Team	Precision	Recall	F1 score
DCU-Gamestreamer	0.0014	0.5709	0.0028
GSE-AAU	0.0112	0.0849	0.0197
Random	0.0667	0.0667	0.0667

We expected few submissions for the second subtask and wanted to employ a qualitative, heuristic evaluation. An expert panel with professionals and researchers from the field of game development, game studies, e-sports, and media sciences should have investigated the submissions and judged them for:

(1) Informative value (i.e., is it a good summary of the gameplay),
(2) Accuracy (i.e., does it reflect the emotional up and downs and the skill of the play), and
(3) Innovation (i.e., surprisingly new approach, non-linearity of the story, creative use of cuts, etc.)

Unfortunately, we did not receive submissions for the second subtask.

5 DISCUSSION AND OUTLOOK

While the MediaEval Emotional Mario task is the spiritual successor of the Gamestory task [5, 6], the goals are different. The work on Counter-Strike: Global Offensive and the analysis of the game streaming and e-sports phenomena have shown the substantial impact games have on culture and society. With the availability of biometric sensors and deep learning for data analysis, we re-focus on the interrelation of the game and the player's experience.

With the Emotional Mario task, we hope to outline the direction of research where player-game interaction can be extended, and games as engines of experience can be understood. Games are not only a playground for people. They are also a vast resource for research and future developments.

ACKNOWLEDGMENTS

We'd like to thank Dr. Andreas Leibetseder for his support.

[2]Evaluation scripts and creation of the random baseline can be found on https://github.com/dermotte/EmotionalMarioEvaluation

REFERENCES

[1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies* 135 (2020), 102370.

[2] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. *arXiv preprint arXiv:1606.01540* (2016).

[3] Justin Shenk et al. 2021. Facial Expression Recognition with a deep neural network as a PyPI package. (2021). https://github.com/justinshenk/fer

[4] Maurizio Garbarino, Matteo Lai, Dan Bender, Rosalind W Picard, and Simone Tognetti. 2014. Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. IEEE, 39–42.

[5] Mathias Lux, Michael Riegler, Duc-Tien Dang-Nguyen, Marcus Larson, Martin Potthast, and Pål Halvorsen. 2018. GameStory Task at MediaEval 2018.. In *Proceedings of MediaEval*.

[6] Mathias Lux, Michael Riegler, Duc-Tien Dang-Nguyen, Johanna Pirker, Martin Potthast, and Pål Halvorsen. 2019. GameStory Task at MediaEval 2019.. In *Proceedings of MediaEval*.

[7] Anvita Saxena, Ashish Khanna, and Deepak Gupta. 2020. Emotion recognition and detection methods: A comprehensive survey. *Journal of Artificial Intelligence and Systems* 2, 1 (2020), 53–79.

[8] Henrik Svoren, Vajira Thambawita, Pål Halvorsen, Petter Jakobsen, Enrique Garcia-Ceja, Farzan Majeed Noori, Hugo L Hammer, Mathias Lux, Michael Alexander Riegler, and Steven Alexander Hicks. 2020. Toadstool: a dataset for training emotional intelligent machines playing Super Mario Bros. In *Proceedings of the ACM Multimedia Systems Conference (MMSYS)*. 309–314.

[9] Tynan Sylvester. 2013. *Designing games: A guide to engineering experiences.* " O'Reilly Media, Inc.".