Multi-Modal Study of the Effect of Time Pressure in a Crisis Management Game

Paris Mavromoustakos-Blom
Tilburg University
p.mavromoustakosblom@uvt.nl

Sander Bakkes
Utrecht University
s.c.j.bakkes@uu.nl

Pieter Spronck
Tilburg University
p.spronck@uvt.nl

ABSTRACT

In this paper, we study the effect of time pressure on player behaviour during a dilemma-based crisis management game. We employ in-game action tracking, physiological sensor data and self-reporting in order to create multi-modal predictive models of player stress responses during a crisis management scenario. We were able to predict the experimental condition (time pressure vs. no time pressure) with 84.5% accuracy, using a game-only feature set. However, lower accuracy was observed when physiological sensor data was used for the same task. The method presented in this paper can be employed in crisis management training, aiming at assessing players' responses to stressful conditions and manipulating player stress levels to provide personalised training scenarios.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; • **Applied computing → Interactive learning environments**.

KEYWORDS

Game-based training, serious games, multi-modal player modeling, crisis management

ACM Reference Format:
Paris Mavromoustakos-Blom, Sander Bakkes, and Pieter Spronck. 2020. Multi-Modal Study of the Effect of Time Pressure in a Crisis Management Game. In *International Conference on the Foundations of Digital Games (FDG '20), September 15–18, 2020, Bugibba, Malta.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3402942.3403006

1 INTRODUCTION

Crisis management is a highly stressful task. Professional crisis responders need to develop a multitude of behavioural competencies including leadership, teamwork, and stress resilience, in order to be able to deal with crises efficiently and consistently. To maintain a high level of preparedness, crisis responders attend regular training sessions where artificial crisis scenarios are "solved" in a cooperative manner. In this paper, we simulate a crisis management training session through the use of an applied game called the Mayor's game [27]. The Mayor's game is a text-based decision making game which allows running dilemma-based crisis scenarios while monitoring player in-game and physiological activity. Our aim is to model player behaviour under the effect of an artificially induced stressor.

To recreate a realistic crisis management setting, we need to be able to increase players' stress levels during gameplay. There are many methods in which player stress levels can be manipulated during a game session, such as time pressure, information complexity, or external distractions. In this study, we selected time pressure as a means of intensifying the game pace and inducing additional stress onto players. Time pressure has been employed as an in-game stress induction mechanism in previous studies revolving around decision making during crises and is considered a condition that is likely to appear in a real-life crisis setting [14]. Therefore, we expect additional time pressure to have a noticeable effect on player physiology and in-game behaviour during a crisis management game session. For the context of the present paper, we refer to the term "stress" as the mental tension players experience when given the responsibility to take complex, high-stake decisions in a limited amount of time.

2 RELATED WORK

Affective computing is defined as computing that relates to, arises from, or influences emotions [23]. Over the years, numerous interfaces to measure, identify and induce human emotions have been employed in human-computer interaction tasks. In this study, we leverage physiological wearable sensors; a non-invasive interface capable of objectively measuring user stress responses.

Physiological sensors have been previously used for user emotion recognition [7, 11], entertainment modeling [16, 30] and stress detection [8, 21]. Physiological signals such as skin conductance and heart activity have been employed as objective descriptors of user player affective state during gaming [16, 28]. More specifically, several studies have focused on the analysis of user experience through biofeedback during shooter games [5, 22]. Our aim is to process and analyse such physiological signals in order to model player behaviour, expecting that artificially induced stress will have an impact on players' physiological measurements.

In this study, we have chosen to employ time pressure as a method of inducing additional stress onto players, aiming to observe the effects on players' physiological responses and in-game behaviour. The influence of time pressure on judgement and decision making has been investigated thoroughly in previous studies focusing both on the cognitive aspects [6, 17, 19], and the practical effects on specific tasks, such as fire ground command [12, 13].

What is particularly relevant to crisis management training, is that the addition of a "deadline" in decision making tasks not only increases humans' feeling of time pressure, but has broader effects

FDG '20, September 15–18, 2020, Bugibba, Malta

ACM ISBN 978-1-4503-8807-8/20/09…$15.00
https://doi.org/10.1145/3402942.3403006

Figure 1: The Mayor's game. Players are required to solve a scenario by answering yes-or-no type dilemmas with the help of five additional advisors.

on their affective state. Through a user study, Maule, Hockey and Bdzola [18] conclude that the induction of time pressure causes increased user awareness, while users were reportedly feeling more "energetic" during a decision making task.

When modeling player stress responses, we anticipate that physiological measurements yield accurate results. We investigate whether additional modalities regarding player behaviour can increase the accuracy of our models. To that end, the present study relies on player modeling based on heartrate (HR) and skin conductance (SC), but also in-game and self-report data. In a relevant study, Holmgård et al. [10] employ blood volume pressure (BVP) and SC signals to implement models of PTSD patients' physiological responses within the StartleMart game. They validate their observations by correlating specific sensor features to submitted player self-reports. Even though treatment efficacy is not investigated through their study, they conclude that their findings can be applied in personalised training environments that support diagnosis and treatment of PTSD. In similar fashion, we explore the possibility of player stress response modeling through multiple modalities, while at the same time investigating possible correlations between them.

In order to build effective crisis management training tools, it is necessary to define the key characteristics that a crisis manager must possess in order to perform when it matters most. Competencies like decisiveness, communication and stress resilience (also called "soft skills" [4]) are only a few examples of abilities that must be developed and trained [25]. "Classical" crisis management training consists of role-playing exercises where crisis scenarios are solved collaboratively [1]. Additional training tools have been introduced, including digital simulations focusing on teamwork & collaboration [24], strategic planning [26] and education of crisis management staff [1, 25]. Large-scale real-life training exercises have also been applied, such as the Bonfire crisis management simulation in the Netherlands [9]. For an extensive review on collaborative crisis management games, we refer readers to Di Loreto et al. [4].

3 METHODOLOGY

In order to study the effects of time pressure in a crisis management game, we have chosen to implement predictive models of player behaviour through multiple modalities. Our models are based on a set of 52 features which were exctracted from in-game logs, physiological sensors and self-reports and have been previously described in [20]. The accuracy of our models along with statistics on feature importance, can interpret the effects that time pressure had on player physiology and playstyle.

We employed the Mayor's game (Figure 1) in order to develop a virtual crisis management scenario. The Mayor's game is a dilemma-based scenario game primarily designed to test leadership skills [27]. Scenarios built for the Mayor's Game have been designed by the Human Behaviour Analytics Lab [15] in close collaboration with crisis management experts. For that reason, and despite the game's low-fidelity design, we expect that through the Mayor's Game we deliver high-quality crisis management experiences.

In the Mayor's game, the player becomes mayor of a fictional town which is undergoing a crisis. The crisis is described by an introductory text and is further unraveled through dilemmas which are sequentially presented to the players. Looking at Figure 1, we see that the main game screen is divided into two parts. In the top half, five advisors are shown, representing experts from various insitutions (police department, fire brigade, legal advisors etc.). The advisors provide additional information to each dilemma in form of text, reflecting their institution's point of view. In order to acquire the additional information, players need to click on the information blobs above the advisors' avatars. Advisors can also provide an answer suggestion on demand, in the form of a green 'tick' (suggesting a 'yes' answer) or a red 'X' (suggesting a 'no' answer). In the bottom half of the screen, the current list of to-be-asnwered dilemmas is presented, while the main text box with the additional information texts, dilemma description and possible answers is shown. Lastly, in the top left of the screen there is a countdown timer, indicating the time left to solve the scenario.

The scenario that is used in our study revolves around a chemical leak resulting from a train accident and consists of eight dilemmas. Each dilemma is accompanied by one piece of additional information from each advisor. Since players need to click on the information blobs to acquire the additional information, reading all the additional information is not mandatory. The dilemmas are all "yes or no" type of questions, and there is not one correct answer; any answer will eventually lead to the end of the scenario.

For this experiment, 82 participants were recruited at Twente University, all current students of the institution. Out of those, 10 participants were excluded because of sensor recording failure. Of the remaining 72 participants, 48 were female and 24 were male, with an average age of 20.48 years ($SD = 1.63$). The experiment was conducted in a laboratory area to minimise distractions from the outside. To collect physiological data, Shimmer3 GSR+ [3] sensors were used. Physiological signal analysis is based on van Gent et al. [29].

Players were divided into two groups. Participants in the control group (N=38) played a baseline version of the scenario lasting 15 minutes, without any external distractions. Participants in the experimental group (N=34), were given three minutes less (12 minutes

Method	Game-only	Sensor-only	Combined
Baseline	52.8% (52.8%)	52.8% (52.8%)	52.8% (52.8%)
Random Forest	84.5% (80.1%)	67.9% (61.1%)	75.0% (68.5%)

Table 1: Results of predicting experimental condition, using game, sensor, or both sets of features combined. Scores outside parentheses represent prediction accuracy when self-report features are excluded from the dataset, while scores in parentheses represent prediction accuracy when self-report features are included.

in total) to solve the same scenario. The amount of time subtracted was heuristically determined, in order to induce time pressure but still allow players enough time to read all the additional information pieces. Additionally, players were reminded of the time left by the experiment coordinators in two minute intervals, while an additional large monitor displaying a countdown timer was placed in the room.

4 RESULTS

This experiment was designed to induce additional stress onto the experimental group's participants through time pressure, aiming to cause variation in physiological and in-game behaviour between participants in the two different groups. The set of features extracted from each modality aims to provide a detailed description of player physiological stress responses and in-game behaviour. We have added a binary variable to our dataset, representing each participant's group (0 for control, 1 for experimental), and ran prediction tasks to investigate whether our extracted feature set can accurately predict that variable. High accuracy in predicting the experimental condition would mean that there is indeed variation in participants' behaviour among the two different groups.

We ran three separate classification tasks, each using a different feature set, accross all participants. We compare two classification methods; as a baseline classifier we used scikit-learn's "dummy classifier", which always predicts the most frequent label in the dataset. We chose to test the accuracy of a random forest classifier (`N_estimators = 10`, `criterion=''gini''`, `max_depth=`∞) through 10-fold cross-validation, which the authors have previously employed successfully for identifying player affective state through facial expression analysis [2].

As illustrated in Table 1, we tested separate predictive models built with in-game and physiological sensor features, as well as a model trained on a combination of both modalities' features. The results show that a random forest classifier achieved 84.5% accuracy when trained on the in-game feature set, while the physiological sensor-based model reached 67.9%. The combination of both feature sets achieved 75% accuracy.

Features *timeToAnswer* and *timeOpen* showed the highest importance during random forest classification. Both of these features are related to players' playstyle (time required to read & answer each dilemma). Participants in the control group tended to spend more time in answering the first four dilemmas of the scenario, leaving them with less time to answer the latter four. On the other hand, participants in the experimental condition tended to distribute time spent more equally accross all dilemmas. This explains the higher accuracy when predictions are made using only the in-game feature set to train the model.

Lastly, we added self-report data including scores of valence and arousal during the game as additional features to the random forest classifier. The reported valence and arousal scores showed no statistically significant difference between the two experimental conditions. Table 1 indicates that adding the self-report features to the dataset, caused a drop in classification accuracy.

5 DISCUSSION

Results indicated that the separate modalities that were used to predict the experimental condition yielded different results. In fact, a model trained using only in-game features, achieved the highest accuracy accross all classification tasks (84.5%). In contrast, models trained using only sensor-derived features, showed relatively low classification accuracy (67.9%). When the two feature sets were unified to train a multi-modal model, we achieved a 75% accuracy.

Since time pressure was increased in the experimental group, we expected the retrieved physiological sensor data to be indicative of that condition. Instead, we notice that the induction of additional time pressure seemed to have more impact on participants' in-game behaviour rather than physiology. Given that the use of the physiological sensor dataset is necessary to generalise this study's results accross multiple games or scenarios, we believe that a multi-modal approach should be preferred. Even with relatively low accuracy, we were able to predict the existence of time pressure regardless of the training scenario currently used. The use of in-game data as the individual input channel to our stress response model greatly increased accuracy; however, without any information about the players' physiological state, we would not be able to explain whether differences in in-game behaviour derive from induced stress or a player's strategical decisions.

In Table 1, we observed a decrease in experimental condition prediction accuracy when self-report features are included in the model. Although we believe that the two conditions were designed with a clear difference in game pace, resulting in high time pressure in the experimental group, one may argue that players cannot reliably report on scenario-wide metrics only at the end of the experiment. While we agree with this statement and believe that a periodical in-game self-report mechanism would yield more reliable results, this was not feasible using the Mayor game's engine.

Lastly, players reported feeling pressured by time during the control (no time pressure) condition, which may lead to the assumption that the baseline scenario duration (15 minutes) already induced a sense of time pressure to the players. As a consequence, both experimental conditions may have been perceived as "high time pressure" by the participants. We believe that a further increase in the difference of time duration between the two conditions may have had a larger impact on the variety of players' stress responses.

6 CONCLUSION

In this study, we collected physiological sensor, in-game and self report data from players during a crisis management scenario game in order to implement models of player behaviour. Our goal was to investigate whether such models can be accurate enough to detect players' physiological and in-game responses to time pressure. In the long term, accurate models may enable the development of realistic personalised crisis management training scenarios, adapted to the individual player through stress level manipulation.

To that end, we ran a classification task: we divided participants in two groups, where participants in the control group played a baseline game, while participants in the experimental group were exposed to artificial stress through time pressure. Subjective measures of players' experienced stress levels were collected through post-game self-reports. We trained models aiming to describe players' physiological and in-game responses to increased time pressure, and implemented a random forest classifier in order to predict the experimental condition. Three separate feature sets were used for classification; a game-only feature set, a physiological sensor feature set, and a combination of both feature sets.

Results show that the two experimental conditions were not perceived as significantly different regaring induced stress levels, as reflected in players' self-reports. However, the induction of stress through time pressure had an impact on physiological measurements, and mostly, in-game behaviour. We were able to predict the current experimental condition with 67.9% accuracy using a random forest model trained on the physiological sensor feature set, while achieving 84.5% accuracy when using the in-game feature set. A model trained using a union of the sensor and in-game feature sets, yielded 75% accuracy. This drop in accuracy may be caused by the random nature of the model's algorithm, where an increase in the number of features can decrease the probability of important features being selected in each iteration. Adding self-report variables to the feature set did not improve classification accuracy.

In conclusion, through an applied crisis management game, we are able to implement accurate predictive models of player behaviour under the effects of time pressure. These results can be employed in future studies in order to implement personalised crisis management training scenarios.

ACKNOWLEDGMENTS

We would like to thank Johannes Steinrucke from Twente University for his assistance during data collection. We would also like to thank the BMS Lab of the University of Twente for providing the sensory equipment. This paper is part of the Data2Game project, partially funded by the Netherlands Research Organisation (NWO).

REFERENCES

[1] Sylvia Babus, Kathryn Hodges, and Erik Kjonnerod. 1997. Simulations and institutional change: Training US government professionals for improved management of complex emergencies abroad. *Journal of Contingencies and Crisis Management* 5, 4 (1997), 231–240.

[2] Paris Mavromoustakos Blom, Sander Bakkes, Chek Tien Tan, Shimon Whiteson, Diederik M Roijers, Roberto Valenti, Theo Gevers, et al. 2014. Towards Personalised Gaming via Facial Expression Recognition.. In *AIIDE*.

[3] Adrian Burns, Barry R Greene, Michael J McGrath, Terrance J O'Shea, Benjamin Kuris, Steven M Ayer, Florin Stroiescu, and Victor Cionca. 2010. SHIMMER™–A wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal* 10, 9 (2010), 1527–1534.

[4] Ines Di Loreto, Simone Mora, and Monica Divitini. 2012. Collaborative serious games for crisis management: an overview. In *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2012 IEEE 21st International Workshop on*. IEEE, 352–357.

[5] Anders Drachen, Lennart E Nacke, Georgios Yannakakis, and Anja Lee Pedersen. 2010. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*. 49–54.

[6] Itiel E Dror, Beth Basola, and Jerome R Busemeyer. 1999. Decision making under time pressure: An independent test of sequential sampling models. *Memory & cognition* 27, 4 (1999), 713–725.

[7] Andreas Haag, Silke Goronzy, Peter Schaich, and Jason Williams. 2004. Emotion recognition using bio-sensors: First steps towards an automatic system. In *Tutorial and research workshop on affective dialogue systems*. Springer, 36–48.

[8] Jennifer A Healey and Rosalind W Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems* 6, 2 (2005), 156–166.

[9] Ira Helsloot. 2005. Bordering on reality: Findings on the bonfire crisis management simulation. *Journal of Contingencies and Crisis Management* 13, 4 (2005), 159–169.

[10] Christoffer Holmgård, Georgios N Yannakakis, Héctor P Martínez, Karen-Inge Karstoft, and Henrik Steen Andersen. 2015. Multimodal ptsd characterization via the startlemart game. *Journal on Multimodal User Interfaces* 9, 1 (2015), 3–15.

[11] Kyung Hwan Kim, Seok Won Bang, and Sang Ryong Kim. 2004. Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing* 42, 3 (2004), 419–427.

[12] Gary A Klein. 2008. Naturalistic decision making. *Human factors* 50, 3 (2008), 456–460.

[13] Gary A Klein, Roberta Calderwood, and Anne Clinton-Cirocco. 1986. Rapid decision making on the fire ground. In *Proceedings of the Human Factors Society Annual Meeting*, Vol. 30. Sage Publications Sage CA: Los Angeles, CA, 576–580.

[14] Gary A Klein and D Klinger. 1991. Naturalistic Decision Making Human Systems. *IAC Gateway* 11, 3 (1991).

[15] Human Behaviour Analytics (HBA) Lab. [n.d.]. ([n. d.]). [Online]. Available: https://www.hba-lab.nl/.

[16] Regan L Mandryk, Kori M Inkpen, and Thomas W Calvert. 2006. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & information technology* 25, 2 (2006), 141–158.

[17] A John Maule and Anne C Edland. 1997. The effects of time pressure on human judgment and decision making. *Decision making: Cognitive models and explanations* (1997), 189–204.

[18] A John Maule, G Robert J Hockey, and Larissa Bdzola. 2000. Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta psychologica* 104, 3 (2000), 283–301.

[19] Ola Svenson A John Maule. 1993. *Time pressure and stress in human judgment and decision making*. Springer Science & Business Media.

[20] Paris Mavromoustakos-Blom, Sander Bakkes, and Pieter Spronck. 2018. Personalized crisis management training on a tablet. In *Proceedings of the 13th International Conference on the Foundations of Digital Games*. ACM, 33.

[21] Paolo Melillo, Marcello Bracale, and Leandro Pecchia. 2011. Nonlinear Heart Rate Variability features for real-life stress detection. Case study: students under stress due to university examination. *Biomedical engineering online* 10, 1 (2011), 96.

[22] Lennart Erik Nacke, Michael Kalyn, Calvin Lough, and Regan Lee Mandryk. 2011. Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 103–112.

[23] Rosalind Wright Picard et al. 1995. Affective computing. (1995).

[24] Aysu Sagun, Dino Bouchlaghem, and Chimay J Anumba. 2009. A scenario-based study on information flow and collaboration patterns in disaster management. *Disasters* 33, 2 (2009), 214–238.

[25] Dirk Stolk, Daniel Alexandrian, Begoña Gros, and Roberto Paggio. 2001. Gaming and multimedia applications for environmental crisis management training. *Computers in Human Behavior* 17, 5-6 (2001), 627–642.

[26] Murray Turoff, Michael Chumer, Xiang Yao, Joseph Konopka, Bartel Van de Walle, and B Carle. 2005. Crisis planning via scenario development gaming. In *Proceedings of the 2nd International ISCRAM Conference, Brussels, Belgium*.

[27] Josine G. M. van de Ven, Hester Stubbé, and Micah Hrehovcsik. 2014. Gaming for Policy Makers: It's Serious!. In *Games and Learning Alliance*, Alessandro De Gloria (Ed.). Springer International Publishing, Cham, 376–382.

[28] Benny van der Vijgh, Robbert-Jan Beun, Maarten Van Rood, and Peter Werkhoven. 2015. Meta-analysis of digital game and study characteristics eliciting physiological stress responses. *Psychophysiology* 52, 8 (2015), 1080–1098.

[29] Paul van Gent, Haneen Farah, Nicole van Nes, and Bart van Arem. 2019. HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation Research Part F: Traffic Psychology and Behaviour* 66 (2019), 368–378.

[30] Georgios N Yannakakis and John Hallam. 2008. Entertainment modeling through physiology in physical play. *International Journal of Human-Computer Studies* 66, 10 (2008), 741–755.