Improving Emotion Detection for Flow Measurement with a High Frame Rate Video based Approach

Ehm Kannegieser[1], Daniel Atorf[1] and Robert Matrai[2]
[1]*Fraunhofer IOSB, Institute of Optronics, System Technologies and Image Exploitation, Fraunhoferstraße 1, 76131 Karlsruhe, Germany*
[2]*FH Technikum Wien, Höchstädtplatz 6, 1200 Wien, Austria*

Keywords: Flow, Immersion, Focus, Emotion Detection, Micro-expressions, High Frame Rate Video.

Abstract: Achieving states of high focus (i.e., Flow, Immersion) in learning situations is linked with the motivation to learn. Developing a tool to measure such states could potentially be used to evaluate and improve learning system potential and thus learning effect. With this purpose in mind, correlations between physiological data and states of high focus were tried to be discovered in a prior study. Physiological data from over 40 participants was recorded and analyzed for correlations with states of high focus. However, no significant correlations between physiological data and elicited states of high focus have been found yet. Revisiting the results, it was concluded that especially the quality and density of emotion recognition data, elicited by a video-based approach might have potentially been insufficient. In this work in progress paper, a method with the intention of improving the quality and density of video data by way of implementing a high frame rate video approach is outlined, thus enabling the search for correlations of physiological data and states of high focus.

1 INTRODUCTION

Serious Games – games, which do not exclusively focus on entertainment value, but rather on achieving learning experiences in players – are successful tools to improve education (Girard, Écalle and Magnan, 2013). In this field of technology, the biggest question is: How might the learning effect be improved even further?

Previous studies have found that the learning effect is linked to the fun games provide to players (Deci and Ryan, 1985; Krapp, 2009), thus, raising the level of fun and measuring its increase becomes more and more important.

Similar to the definition of Flow – as the state of optimal enjoyment of an activity (Csikszentmihalyi, 1991) and Immersion as the sub-optimal state of an experience (Cairns, 2006), fun is defined as the process of becoming voluntarily engrossed in an activity. As such, measuring these states of high focus becomes interesting when analyzing the fun experienced during gameplay (Beume et al., 2008).

Both Flow and Immersion are currently measured using questionnaires (Nordin, Denisova and Cairns, 2014). The questionnaires can be elicited either during the game – disrupting the player's concentration – or after the game, leading to imprecise results. Additionally, questionnaires can only elicit subjective measurements, further degrading the quality of the data gathered.

For this reason, the development of a system for automatic measurement of Immersion and Flow becomes increasingly interesting. Instead of using questionnaires filled out by participants, in previous work (Atorf et al., 2016; Kannegieser et al., 2018) a study was introduced, which aimed to further the understanding of how Flow and Immersion are linked together and to ease future work towards a new measurement method using physiological data to determine their current Flow or Immersion state.

Such a measurement method would provide Serious Game developers with better, more objective insight about how much fun, and respectively, how much learning value is provided by their games.

In a previous study, first steps had been taken in in the direction of developing such a method. Based on the measured data, no correlations between states of high focus and physiological signals were found. However, this does not yet prove that no such

Kannegieser, E., Atorf, D. and Matrai, R.
Improving Emotion Detection for Flow Measurement with a High Frame Rate Video based Approach.
DOI: 10.5220/0009795004900498
In *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020) - Volume 2*, pages 490-498
ISBN: 978-989-758-417-6

correlation exists and better data quality and density might deliver different results.

The next Chapter explores relevant concepts by reviewing related research and outlines the study preceding this one, in which no correlations had been found. Then, Chapter 3 delineates a new measurement method that to be integrated into the existing experiment setup with the intention of improving data quality and density, as suggested above.

2 RELATED WORK

Flow was first described by Csikszentmihalyi as the state of the optimal experience of an activity (Csikszentmihalyi, 1991). When entering a state of Flow, even taxing activities like work no longer feel taxing, but rather feel enjoyable. However, the Flow state cannot be achieved for every activity. Csikszentmihalyi bases flow on the model of extrinsic and intrinsic motivation. Only intrinsically motivated actions, which are not motivated by external factors, can reach the Flow state. Flow is reachable when the challenge presented by such an intrinsically motivated action is balanced with the skill of the person performing the task. All this makes Flow an interesting point of research concerning games, as playing games is usually intrinsically motivated. Flow is mapped to games in the GameFlow questionnaire (Sweetser et al., 2005).

There exist two concurrent definitions of Immersion. The first definition is called presence-based Immersion and refers to the feeling of being physically present in a virtual location. The second definition is known as engagement-based Immersion. It defines Immersion based on the strength of a player's interaction with the game. The model given by Cairns et al. in their series of papers (Cairns et al., 2006; Jenett et al., 2008), defines Immersion as a hierarchical structure, with different barriers of entry. The lowest level, Engagement, is reached by interacting with the game and spending time with it. Engrossment is reached by becoming emotionally involved with the game. During this state, feelings of temporal and spatial dissociation are starting to appear. The final state, Total Immersion, is reached by players having their feelings completely focused on the game. Cheng et al. improved upon this hierarchical model by adding dimensions to the three levels of the hierarchy (Cheng et al., 2015). The Engagement level is split into the three dimensions: Attraction, Time Investment and Usability. The second level, Engrossment, is split into Emotional Attachment, which refers to attachment to the game itself, and Decreased Perceptions. Finally, Total Immersion is defined by the terms Presence and Empathy.

Table 1: Comparison of similarities in Flow and Immersion definitions.

Flow	Immersion
Task	The Game
Concentration	Cognitive Involvement
Skill/Challenge Balance	Challenge
Sense of Control	Control
Clear Goals	Emotional Involvement
Immediate Feedback	-
Reduced Sense of Self and of Time	Real World Dissociation

Flow and Immersion share many similarities. Both have similar effects, such as decreased perceptions of both time and the environment, and refer to a state of focus (see Table 1.).

Georgiou and Kyza even take the empathy dimension in the immersion model by Cheng et al. and replace it with Flow (Georgiou and Kyza, 2017).

There are two main differences between the two definitions: First, Flow does not define an emotional component, while Immersion is focused heavily on the emotional attachment of players to the game. Second, while Flow refers to a final state of complete concentration, Immersion refers to a range of experiences, ranging from minimal engagement to complete focus on the game.

The model used in previous work (Kannegieser et al., 2018) is based on the Flow model presented by Csikszentmihalyi (Csikszentmihalyi, 1991) and the Immersion model by Cheng et al. (Cheng et al., 2015), which itself is a refinement of the hierarchical model presented by Cairns et al.. Flow, as the optimal experience of an action, is considered the highest point in the Immersion hierarchy, which implies that Total Immersion and Flow are regularly experienced together. Figure 1 presents the Immersion hierarchy imposed on top of the three-channel model by Csikszentmihalyi. As Immersion grows the possibility to reach the Flow state increases. It must be noted that the diagram is only meant to be a qualitative visualization, as Immersion is not dependent on the challenge/skill balance.

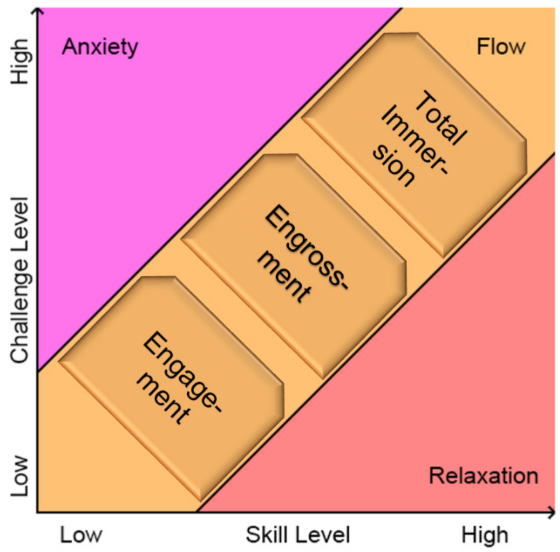

Figure 1: Combined model of Flow and Immersion. Qualitative view, the skill/challenge balance does not influence Immersion.

2.1 Experiment

The mentioned study of Kannegieser et al. (Kannegieser et al., 2019) was designed to both gather data to link physiological measurements with Flow and Immersion, as well as validate the Flow/Immersion model presented in section 2.

About forty participants took part in the study. The number of participants chosen for the experiment is similar to the number of participants used in other experiments in this area (Cairns et al. 2006, Jennett et al., 2008). There were no requirements for participants, which were self-selected as the experiment was aiming for as close to a random selection as possible and to observe higher levels of Immersion and Flow.

The study was split into three phases. During the Setup Phase, a game was selected. Free choice of game makes finding links between physiological measurements harder, but was chosen to help participants reaching the Flow state more easily. During the Gaming Phase, participants played the game for 30 minutes. After the Gaming Phase had concluded, participants entered the Assessment Phase and watched their previous gaming session, while answering questionnaires about Immersion and Flow periodically. This setup was chosen to get more exact results and because it does not interrupt the Flow experience.

Three questionnaires were used during the study. The first questionnaire used was the Immersion questionnaire described by Cheng et al (Cheng et al., 2015). As the questionnaire was too long to be measured multiple times without worsening the results, it was split into an Immersive Tendency questionnaire asked at the beginning of playback and an iterative questionnaire asked every three minutes during playback. For Flow, the Flow Short Scale questionnaire by Rheinberg et al. was used (Rheinberg et al., 2003). It was originally designed for being used multiple times in a row, making it perfect for this iterative approach. During playback, it is asked every six minutes. The final questionnaire used is the Game Experience Questionnaire (IJsselsteijn, de Kort and Poels, 2013). It measures a more general set of questions and was asked once after playback is over. Figure 2 shows the three phases of the experiment as well as the activities of each phase.

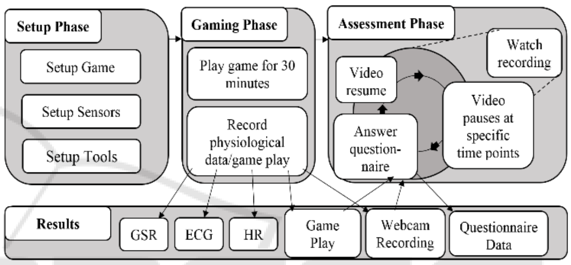

Figure 2: Three phases of the experiment.

2.1.1 Physiological Measurements

For 30 minutes, physiological measurements were taken. The physiological measurements used in this study were used due to being non-intrusive and not hindering the immersion of players. Galvanic skin response, Electrocardiography, gaze tracking and web cam footage for emotion analysis were used. A facial EMG would be more precise for analyzing displayed emotions, however, placing electrodes on the face of the player would distract from the game play experience and make it harder to reach the flow state. For the same reasons, sensors for EEG measurement were not chosen for the study.

Aside from Galvanic Skin Response (GSR), Electrocardiography (ECG), Eye-tracking and screen game play recordings, web cam footage of the player and was obtained with a resolution of 960x720 and a frame rate of 15 fps. From this web cam footage, the facial portion of the still images were selected and emotion recognition was performed on this extraction Convolutional Neural Networks (CNN) with the method proposed by Levi and Hassner (Levi and Hassner, 2015).

2.1.2 Analysis and Results

In the first step of the analysis, the data was checked for correlations between Flow and Immersion. As the results from both the Flow and Immersion questionnaires did not follow a normal distribution, Spearman correlation was chosen. The correlation analysis found a strong correlation between all three levels of Immersion and Flow. The strongest correlation was found between Engagement and Flow (R = 0.69, p = 8.536e-30), which made sense, knowing that Flow encompasses all features making up Engagement. The second strongest correlation exists between Total Immersion and Flow (R = 0.652, p = 1.91e-25) (see Figure 3). This is caused by the fact that players who played games without clear avatars, such as strategy games, found it difficult to emphasize with their avatar in the game, leading to reduced Total Immersion. The least correlated level of the three was Engrossment (R = 0.56, p = 1.829e-18), which can be explained as Engrossment puts strong emphasis on emotional attachment of the player to the game, something Flow does not elicit. All three showed strong correlation to Flow (see Table 2), meaning the relation between these two psychological states explained in section 2 is likely.

Table 2: Correlation between Flow and Immersion (Spearman-Rho-Coefficient).

	Flow	Engage-ment	Engross-ment	Total Immersion
Flow	1	0.69	0.57	0.65
Engage-ment	0.69	1	0.45	0.58
Engross-ment	0.57	0.45	1	0.62
Total Immersion	0.65	0.58	0.62	1

Table 3: Correlation between Flow and physiological measurements (Spearman-Rho-Coefficient).

	GSR	HR	Fixations per minute
Flow	-0.02	-0.03	-0.07
Engage-ment	0.01	-0.08	-0.02
Engross-ment	-0.04	-0.09	0.05
Total Immersion	-0.15	0	0.06

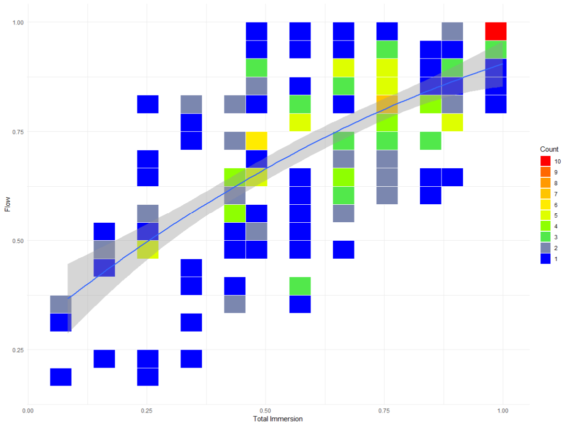

Figure 3: Scatter plot for correlation between Flow and Total Immersion, R=0,65; P<2,2e^-16; conf=0,95.

Direct correlation between normalized physiological data and answers of the Flow and Immersion questionnaires showed no meaningful correlation. The direct correlation results are shown in Table 3. Further discussion on the statistical methods employed can be found in Kannegieser et al., 2019.

3 ONGOING WORK

Given the similarities in definition, a correlation between Engagement-based Immersion and Flow seemed a logical consequence, as shown in the previous chapter. However, the data elicited by a study to find the link between physiology and high focus states did not yield matching results.
Therefore, ongoing work focuses on expanding the study setup and refining the methods utilized to gain detailed insight and improve the understanding of the data recorded as well as improve overall data quality and density. How to elicit Micro-expressions (ME) which are defined as "true emotional state" which are deemed suitable to help finding more significant relations between states of high focus and physiological data, is elaborated on further in this chapter.

3.1 Micro-expressions

ME are very short (40-200ms) contractions of facial muscles in limited areas (Liong et al., 2019; Gan et al., 2019). ME contrast macro-expressions (MA) (>200ms) in duration, intensity, and the scope of the affected areas.

Unlike MA, ME are also involuntary in nature, i.e. they emerge without conscious intent and cannot be replicated deliberately. That is, micro-expressions are not subject to conscious manipulation and thence reflect peoples' true emotions.

Capturing and identifying ME with adequate hard- and software could enable the inference of emotional states experienced by participants. The ability to detect emotions in this manner could prove to be a valuable addition to our experiment setup in the context of measuring high focus states.

3.2 Capturing Micro-expressions

As in the case of macro-expressions, there are two established methods to capture micro-expressions (Ekman, 1992; Tan et al. 2012). The first method involves measuring the activity of facial muscles using electromyography (EMG). The second method employs special software to detect facial expressions visually based on footage acquired by video cameras.

Recording ME via electromyography is performed by measuring several facial muscle regions of the mimetic muscles (Fridlund and Cacioppo, 1986). As the placement of electrodes onto the face of the participant had been deemed too invasive, High Frame Rate Video (HFRV) was chosen as an alternative approach.

Conventional video cameras record video with either 30 frames per second (fps) (NTSC – e.g. in North America), 25 fps (PAL – e.g. in Europe) or 24 fps (cinema). Although the term is not defined precisely, HFRV is understood to refer to video with frame rates higher than these conventional frame rates.

ME will be identified by first segmenting the captured videos into individual pictures, then extracting the facial area by a machine learning (ML) algorithm, and finally detecting emotions using the CNN by Levi and Hassner (Levi and Hassner, 2015).

3.3 Feasibility Study of HFRV

The experiment at hand is intended to accumulate data with the goal of determining connections between physiological signals and immersive states. This automatically poses the requirement on all sensors and electronics used, that these not impede participants from experiencing said immersive states. From the two methods for capturing ME, HFRV has been selected for integration into the existing study setup, due to its less invasive nature in comparison to facial EMG.

3.3.1 HFRV with the Current Setup

In the current study setup, video footage of the participants is acquired using the web cam Logitech C920 (Logitech International S.A., Newark, California, USA). Theoretically, this camera is capable of recording video with a maximum of 30 fps. With all other software running on the computer at the same time, the highest achieved sampling rate, without overall negative performance impact was 15 fps.

Most cameras are controlled by internal electronics, and save recordings to a storage medium, like a memory card. Web cams, on the other hand, can be controlled by software on the computer they are attached to and save the videos directly to the hard drive of the computer, which adds additional load to the computer.

Video games can have high memory requirements, as does the software used for recording multiple physiological signals and screen capturing. The combination of all these processes resulted in disturbances to the player in the form of slowdowns, reductions in the game's frame rate and buffer issues when writing data streams to disk: the frame drop of the screen capture and web cam video increased significantly, when ramping up the video output frame rate, resulting in deteriorated data quality.

In order to quantify the effects of capturing video via web cam on the computer's performance, multiple benchmark tests have been carried out. First, without recording video, then with ever-increasing frame rates. Two different benchmarks were used: the 3DMark basic edition (Futuremark Oy, Espoo, Finland), and the Unigine Heaven Benchmark 4.0 (Unigine Corp., Clemency, Luxembourg). In all three tests, the achieved overall scores showed negative tendencies, indicating that recording video on the web cam mentioned above has a negative impact on the computer's performance. The exact results can be seen in Table 4. The numerical values show that increases in frame rates cause only small changes in performance. However, the differences in subjectively perceivable spatial resolution during the tests were substantial.

One way to resolve the resulting bottleneck regarding the computer's performance would be upgrading the computer. This solution has been repudiated, as it would have required substantial financial investment. As an alternative, it has been proposed that the current web cam be replaced with a different camera. Potentially, this could prevent disturbances to immersion in the game, while also capturing video at higher, more adequate frame rates.

Table 4: Overall scores of the computer used in the experiment on three benchmark tests with different frame rates.

FPS	3DMark		Heaven	
	Score	%CPU	Score	%CPU
No video	10484	91.707	2685	93.768
10	10768	94.593	2687	97.581
15	10678	95.443	2586	94.196
20	10532	94.911	2587	96.821
25	10451	98.489	2580	99.651
30	10332	99.314	2566	98.946

As reported in the scientific literature, the shortest ME last a mere 40ms, or 1/25 of a second. Theoretically, to capture each signal, each ME, a sampling rate higher than the minimal frequency of the original signal should be chosen. Therefore, in the case of this experiment, a minimum 26 fps is necessary.

Using even higher frame rates would insure that each signal is captured with higher certainty, while also providing additional information in regards to each individual ME. In this manner, information pertaining to the path of the movement could be acquired, as well, potentially improving the accuracy of emotion detection.

Common frame rates of camera hardware, able to record video faster than 30 fps include 50, 60, 90, 120 and 240 fps. To allow a detailed sampling of the target signal and to coincide with a conventional frame rate, the minimum necessary frame rate for this selection process has been set at 60 fps.

Three cameras available in-house, the Sony HDR-CX240E, the Sony Alpha 5100, and the Sony FCB-ER8550 (Sony Corporation, Minato, Tokyo, Japan), have been tested. The cameras were operated via a USB-HDMI-Interface, an Elgato HD-Cam Link (Corsair GmbH, Munich, Germany) with an internal restriction to 60 fps), and the maximum possible frame rate has been assessed. Each of these cameras achieved maximum frame rates higher than the aforementioned webcam. The exact results can be seen in Table 5.

Table 5: Evaluated cameras and the respective frame rates achieved.

Camera	Achieved FPS
Logitech C920	29
Sony HDR-CX240E	30
Sony Alpha 5100	50
Sony FCB-ER8550	59

As these cameras could not achieve frame rates of 60 fps, alternative camera equipment available on the market has been selected based on the requirements listed in 3.3.2. To evaluate their suitability for emotion recognition, the selected cameras will be integrated into the experiment setup and tested. Videos captured by each camera will be evaluated regarding their performance in ME and emotion recognition with multiple frame rate settings (60, 120, 240 fps).

3.3.2 Requirements

In order to be classified as eligible for integration into the experimental setup, cameras should meet certain requirements regarding hardware features. As they greatly simplify and shorten developmental processes, integrating control functions for the camera into the existing software with the help of an application-programming interface (API) is required to be feasible.

High Resolution. Spatial resolution of the camera should be high enough to give detailed visual information of the subject's face. High resolution would also allow for placing the camera far enough from the subject to provide them with a certain level of freedom of movement. Potentially, this could make participants more at ease and promote immersion. Full-HD (1080p) had been set as a target value.

High Frame Rate. As outlined above (3.3.1), the minimum frame rate has been set at 60 fps. However, using even higher frame rates could deliver more detailed information regarding facial muscle movements.

Instead of using the pre-trained CNN of Levi and Hassner (2015) for face detection and extraction, it would also be conceivable to train this CNN with self-generated data, or to use a different network, also self-trained. In this manner, face detection accuracy could potentially be improved. For training a neural network, a large training data set is essential. Higher frame rates could provide more data per measurement, contributing to and facilitating the accumulation of such a data set (Pfister et al., 2011).

Internal Soft- and Hardware for Video Recording and Storage. These features would allow for the outsourcing of the computational burden of capturing and saving video. Outsourcing these tasks would free up computational capacity on the main computer, contributing to a lag-free gaming experience.

API. An available open source API would be rather advantageous, because it would greatly simplify the integration of the camera into the existing experimental setup and software. In addition, this would do away with the restrictions the HD-Cam Link poses on frame rates (60 fps). With the help of said API, the following functionalities should be feasible: Set camera settings (Resolution, FPS, FOV, etc.), Start/stop video recording, Media export to computer and deleting media from memory card.

3.3.3 Proposed Solution

For subsequent integration into our experiment setup, off-the-shelf cameras on the market have been evaluated, based on the criteria listed above (3.3.2).

Two action cameras (action-cams) have been selected and purchased for further testing: the Yi4k (Xiaoyi Technology Co., Ltd., Pudong District, Shanghai, China) and the GoPro Hero7 Black (GoPro Inc., San Mateo, California, USA). The maximum frame rate of the Yi4k is 120 fps, and that of the Hero7 Black is 240 fps, both at a resolution of 1080p. At this resolution, both cameras are able to record at their respective highest frame rates. Both are also capable of recording at higher resolutions, albeit only at lower frame rates.

Both cameras use internal soft- and hardware for capturing and storing video. Moreover, these cameras can be controlled via API. In theory, this should allow for outsourcing the computational burden of capturing videos while synchronizing said videos with the recorded physiological signals. That is, these two action-cams meet all four requirements mentioned above.

For the Yi4k, there is an open-source API (Yi Technology, 2017) available online. To the contrary, the GoPro Hero7 Black has none. Fortunately, however, it can be controlled via simple HTTP-requests. A list of these requests is available online (Iturbe, 2020). Utilizing said API, both cameras will be integrated into the current setup: the video recordings will be started and stopped and file transfer over Wi-Fi will be initiated from the computer.

Multiple tests will be carried out regarding these cameras: some regarding the performance of video acquired by these cameras with different frame rates (60-240 fps) in ME and emotion recognition, and others regarding circumstantial modalities. These circumstantial modalities include battery life, speed of data transfer over Wi-Fi and its effect on experiment duration, and utilized color encoding systems. It is imperative that the battery last long enough to record one experiment and carry through data transfer to the computer. In this experiment, the recorded videos are 30 min long. HFRV-files of this length will be several GB in size. Therefore, the length of time necessary for data transfer to the computer over Wi-Fi will have to be assessed. With the color encoding system NTSC, higher frame rates can be achieved than with PAL. Therefore, NTSC would be preferred. The compatibility of this setting with artificial lighting under European standard AC frequency will have to be tested, as well.

4 CONCLUSIONS AND DISCUSSION

This paper gave an overview of the current state of work related to the physiology of Flow and Immersion. It referenced a preceding study that did not yield the expected results, but also did not rule out the possibility of achieving such results with different methods. It laid out the experiment setup of the previous study and delineated plans to expand it with the goal of improving data quality and density. This refers to the integration of a HFRV approach.

With the integration of the proposed method, further insight regarding the relationship between Flow/Immersion and physiological signals could be gained. Obtaining such insight could prove to be a step forward in developing a tool for measuring high focus states physiologically.

Further plans have been described to boost the performance of machine learning methods already in use (face detection/extraction, emotion recognition) as well as to employ machine learning as a substitute for conventional statistical analysis in identifying relationships between physiological signals and questionnaire data.

Currently, both face detection/extraction and emotion recognition is accomplished using software developed and described by Levi and Hassner (Levi and Hassner, 2015). This program, like any other, is not fully accurate. It does not recognize faces in images with perfect precision, misidentifications are inevitable.

The software used for emotion recognition faces a similar problem: As detailed in their paper, the CNN of Levi and Hassner used for emotion recognition accurately classified approximately 54% of the displayed emotions into seven categories. In the concrete application laid out in this work, this accuracy is to be improved by better quality footage. However, improving video quality is not the only way imaginable to achieve such improvement. One possibility would be to train the aforementioned CNN with self-generated and application-specific data. As this method runs into the difficulty of labelling data, other methods seem more actionable. For example, alternative NN could provide better results. As of 2020, the CNN used in this work is about five years old; it seems plausible to think that in the rapidly evolving field of machine learning other NN with higher classification accuracy have been developed in the meantime.

Video data is not the only kind of data collected. Parallel to capturing video footage, other measurement systems are also in use. These include EMG, ECG, and GSR (Kannegieser et al., 2018). In these cases, similar to video data, there could still be room for improvement regarding data quality, as well. Such improvements could theoretically be achieved using alternative measurement tools or different methods for data processing.

Up until now, statistical methods have been used for finding correlations between questionnaire data and physiological signals. As mentioned before in this paper, none has been found. Apart from improving the quality of the data with methods like the ones described above, one could entertain the idea that such correlations could be found with different analytical methods. For example, as a tool capable of establishing connections based on high-level abstraction, machine learning seems an obvious and promising candidate.

REFERENCES

Atorf, D., Hensler L., and Kannegieser, E. (2016). Towards a concept on measuring the Flow state during gameplay of serious games. In: *European Conference on Games Based Learning (ECGBL)*. ECGBL 2016. Paisley, Scotland, pp. 955–959. isbn: 978-1-911218-09-8.

Beume, N. et al. (June 2008). “Measuring Flow as concept for detecting game fun in the Pac-Man game”. In: *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pp. 3448–3455. doi: 10.1109/CEC.2008.4631264.

Cairns, P. (2006). “Quantifying the experience of immersion in games”.

Cheng, M.-T., H.-C. She, and L.A. Annetta (June 2015). “Game Immersion Experience: Its Hierarchical Structure and Impact on Game-based Science Learning”. In: *Journal of Computer Assisted Learning 31.3*, pp. 232–253. issn: 0266-4909. doi: 10.1111/jcal.12066.

Csikszentmihalyi, M., (Mar. 1991). Flow: The Psychology of Optimal Experience. New York, NY: Harper Perennial. isbn: 0060920432.

Deci, E. and Ryan, R. (Jan. 1985). Intrinsic Motivation and Self-Determination in Human Behavior. Vol. 3.

Ekman, P. (1992). Facial expressions of emotion: an old controversy and new findings. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 335(1273)*, 63-69.

Fridlund, A. and Cacioppo, J. (1986). Guidelines for Human Electromyographic Research. In: *Psychophysiology. 23*. 567 - 589. 10.1111/j.1469-8986.1986.tb00676.x.

Gan, Y., Liong, S.-T., Yau, W.-C., Huang, Y.-C. & Tan, L.-K., 2019. Off-apexnet on microexpression recognition system. In: *Signal Processing: Image Communication, 74*, pp.129–139.

Georgiou, Y. and Kyza, E.-A. (Feb. 2017). “The Development and Validation of the ARI Questionnaire”. In: *International Journal of Human-Computer Studies 98.C*, pp. 24–37. issn: 1071-5819. doi: 10.1016/j.ijhcs.2016.09.014.

Girard, C., Écalle, Jean, and Magnan, Annie, (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. In: *Journal of Computer Assisted Learning 29*, pp. 207–219.

IJsselsteijn, W. A., de Kort, Y. A. W., and Poels, K. (2013). The Game Experience Questionnaire. Eindhoven: Technische Universiteit Eindhoven.

Iturbe, K., goprowifihack (2020). Available at: https://github.com/KonradIT/goprowifihack [Accessed 05.02.2020].

Jennett, C. et al. (Sept. 2008). “Measuring and Defining the Experience of Immersion in Games”. In: *Int. J. Hum-Comput. Stud. 66.9*, pp. 641–661. issn: 1071-5819. doi: 10. 1016/j.ijhcs.2008.04.004.

Kannegieser, E., Atorf, D., and Meier, J. (2018). Surveying games with a combined model of Immersion and Flow. In: *MCCSIS 2018 Multi Conference on Computer Science and Information Systems, Game and Entertainment Technologies*. 2018.

Kannegieser, E.; Atorf, D. and Meier, J. (2019). Conducting an Experiment for Validating the Combined Model of Immersion and Flow. In *Proceedings of the 11th International Conference on Computer Supported Education - Volume 2: CSEDU*, ISBN 978-989-758-367-4, pages 252-259. DOI: 10.5220/ 0007688902520259

Krapp, A., Schiefele, U., and Schreyer, I., (2009). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. postprint.

Levi, G., and Hassner, T., (2015). Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns. In: *Proceedings ACM International Conference on Multimodal Interaction (ICMI)*, Seattle, 2015.

Liong, S.-T., Gan, Y., See, J., Khor, H.-Q, and Huang, Y.-C., 2019. Shallow triple stream threedimensional cnn (ststnet) for micro-expression recognition. In: *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pp.1–5.

Nordin, A., Denisova, A. and Cairns, P. (2014). Too many questionnaires: measuring player experience whilst playing digital games. In *Seventh York Doctoral*

Symposium on Computer Science & Electronics, pp.69-75.

Pfister, T., Li, X., Zhao, G. & Pietikäinen, M., 2011. Recognising spontaneous facial microexpressions. In: *2011 international conference on computer vision*, pp.1449–1456.

Rheinberg, F., Vollmeyer, R., and Engeser, S., (2003). “Die Erfassung des Flow-Erlebens”. In: *Diagnostik von Motivation und Selbstkonzept*. Göttingen: Hogrefe, pp. 261–279.

Sweetser, P. and Wyeth, P. (July 2005). “GameFlow: A Model for Evaluating Player Enjoyment in Games”. In: *Comput. Entertain. 3.3*, pp. 3–3. issn: 1544-3574. doi: 10.1145/1077246.1077253.

Tan, C. T., Rosser, D., Bakkes, S., and Pisan, Y. (2012). A feasibility study in using facial expressions analysis to evaluate player experiences. In: *Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing the System* (pp. 1-10).

YI Technology, I., 2017. YIOpenAPI Available at: https://github.com/YITechnology/YIOpenAPI [Accessed 05.02.2020].