Flow Adaptation in Serious Games for Health

Tomás Alves
Department of Computer Engineering
Instituto Superior Técnico
Lisboa, Portugal
tomas.alves@tecnico.ulisboa.pt

Sandra Gama
Department of Computer Engineering
Instituto Superior Técnico
Lisboa, Portugal
sandra.gama@tecnico.ulisboa.pt

Francisco S. Melo
Department of Computer Engineering
Instituto Superior Técnico
Lisboa, Portugal
fmelo@inesc-id.pt

Abstract—This work studies the relevance of flow in gameplay adaptability and how it may offer a better gaming experience, since it provides a better enjoyment of an activity. We developed a first-person shooter video game that adapts its in-game difficulty and environmental settings based on a representation of the mental state of the user to keep a balance between the skills of the player and the challenge of the game. The mental state of the player is measured with their physiological signals, namely the heart rate and the beta band of the brainwaves, and we distinguish the mental state of the player with an accuracy of 87%. We also conducted an evaluation using self-perceived flow and in-game scores as metrics to compare the mental state-based adaptability with a performance-based version. Results show that the latter provided a better gaming experience, suggesting that further research is needed to fully understand the relation of flow and gameplay on FPS games.

Index Terms—Videogames, flow, performance, adaptable gameplay, psychophysiology

I. Introduction

Video games are part of our culture, not only for entertainment, but also to educate and train people [1]. There are already studies using video games with patients experiencing diseases or physical disabilities [2]–[4]. More specifically, Schlickum *et al.* [5] examined the causal effect of a First-Person Shooter (FPS) on virtual surgical endoscopy skills and results suggest that the content and demands of video games are important for a transfer to surgical skills to occur.

Naturally, the greater the motivation to participate in rehabilitation tasks, the more likely will be to patients continuing an active participation. Therefore, if we can provide a greater motivation to play, patients will play for a longer time. It is known that longer therapeutic sessions lead to greater functional outcomes over the course of treatment, therefore games should keep players interested in playing until they are healthy again. Several games [6], [7], already try to adapt their gameplay to the patient in order to provide a better experience.

The mental state associated to the optimal enjoyment of an activity is flow. This concept is associated to the mental state that people feel when they are completely engaged in an activity and have an optimal experience while performing it [8]. The balance between the challenge of an activity and the skills of an individual is one of the conditions that lead people to flow. A game that adapts its gameplay difficulty and environment interactions allows a better gaming experience compared to a game in discrete difficulties, using the gameplay

difficulty to balance the challenge of the game and the skills of the player and the environment interactions to create a greater sense of engagement in the player [9]. Therefore, **since flow provides a better enjoyment of an experience [8], we study the relevance of flow in gameplay adaptability and if it may offer an enhanced gaming experience**.

Our hypothesis is that **gamers have a better gaming experience playing a game that adapts to the representation of their mental state by keeping them in flow compared to a game that adapts to their performance**. To test our hypothesis, we developed two prototypes in user testing to evaluate self-perceived flow and in-game performance: one adapts to the mental state of the user and the other to their performance. As a metric for success, we defined that players have a higher performance and a higher self-perceived flow when they are playing a game that adapts to their mental state compared to playing a game which gameplay adapts to their performance.

II. Background

Flow regarding human behavior and computers has most notably been studied as a means to explain user engagement. This section presents the flow theory and how we can physically measure the flow state.

II-A. Flow

Csikszentmihalyi [10] addresses the feeling of deep engagement as state of "flow". Based on his findings, Csikszentmihalyi defined the two conditions for flow: (1) Perceived challenges of the activity match and stretch the capabilities of the individual, thus producing an experience of being fully engaged in the task and acting on the height of their skills [11], [12]; and (2) The goals of the activity are explicit and reachable, and one receives instant feedback for their progress on the activity [12].

Nacke and Lindley [13] presented a two-dimensional four-channel model of flow based on Csikszentmihalyi [8] and Ellis *et al.* [14] which incorporates the apathy state and is used most frequently for describing games and gameplay experience. We based our adaptability in this model and it is illustrated in Figure 1.

II-B. Flow in Experimental Settings

The first-level physiological indicators that are proven to be effective measuring flow are the Heart Rate (HR) and Heart
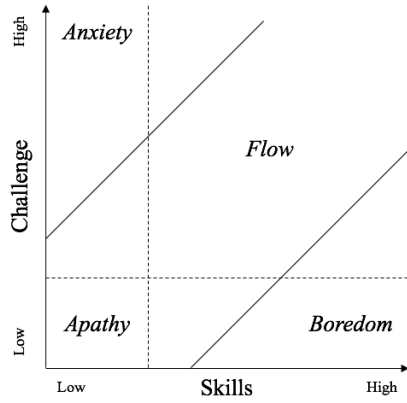
Fig. 1: Four-channel flow model presented by Nacke and Lindley [13]. The flow state (an equilibrium between skills and challenge) is compared against anxiety (challenge exceeds skills) and boredom (skills exceed challenge) conditions. Apathy was reported when challenges and skills were too low at the start or when a task had to be repeated frequently.

Rate Variability (HRV) [15], [16] for cardiovascular activity, the Respiratory Rate (RR) [15], [16] for respiratory activity and the alpha, low beta and theta bands [17], [18] for the Electroencephalography (EEG). The decrease of the Galvanic Skin Response (GSR) is also an indicator for higher arousal and can be effectively used to measure it [19], [20].

In particular, we seek to detect different representations of the mental state of the user using these signals. In order to do so, we place sensors on the users while they are playing the different prototypes and record their biofeedback. These signals were also selected because they are rather resistant to movement artifacts and can be measured non-invasively. More details are discussed in section III-B1.

III. ADAPTABLE GAMEPLAY DEVELOPMENT

In order to create a game which gameplay adapts to the user's mental state, we need three different components. First, the game itself which must implement a gameplay adaptation mechanism. With this mechanism, it is possible to change the game parameters while the user is playing the game. Second, it is necessary to know the mental state of the player so that we can take a proper adaptation. We need a classification framework to read the physiological signals of the player and label the output as the current mental state of the player. Third, by knowing the current mental state of the player and taking into account the Flow Theory [8], we can choose the difficulty by changing the game parameters to match the skills of the player and keep them in a flow state: if the player is anxious, we decrease the difficulty; and, if the player is bored, we increase difficulty. The controller framework is responsible for reading the current mental state of the player and varying the game parameters, depending on the previous ones.

III-A. Game

The game is a FPS in a dark environment with cartoonish zombie figures as enemies, based on the Unity3D "Survival

Shooter tutorial"[1], as illustrated in Figure 2.

Fig. 2: Representative picture of the Nightmares Survival Shooter tutorial in Unity Asset Store.

We choose a first-person camera because it will easily lead to user immersion [21]. Immersion involves a loss of a sense of context, while flow describes a level of complete involvement [11]. Immersion in a FPS is possible given the removal of the avatar representation and putting the player in first-person perspective, which leads the player to feel like they are acting directly upon the virtual game world [22]. Thus, the player can fully identify with the game character represented only by the weapons that reach into the game environment. Takatalo *et al.* [23] and Nacke and Lindley [13] also used a FPS game to study user engagement.

The zombie genre gained popularity in recent years due to the development of themes like *The Walking Dead*[2], *Resident Evil*[3] and *Left 4 Dead*[4]. Being an endless shooter game, this work relies on the survival challenge and familiarity associated to the zombie genre to promote user engagement. This section covers the base design and the development of mental state inducement versions.

III-A1 Base Game

The game comprises three different enemies: Zombunnies (low health points and high speed), Zombears (medium health points and medium speed) and Hellephants (high health points and low speed). We introduced several new features based on the work of Cowley *et al.* [24]. Since they mapped flow and gameplay elements according to Csikszentmihalyi's dimensions of flow [11], we decided to create gameplay elements that are supported by the mapping. The new features include the first-person camera, a rocket launcher, stamina and sprinting system and a pick-up system with rewards of health and ammunition that the player can catch scattered across the game world.

III-A2 Inducement Versions

A key point in this work is that our contribution must be able to distinguish the mental state of the user. Therefore, we need to have data of the different mental states. We created four different versions of the base game by varying some of

---

[1] www.unity3d.com/learn/tutorials/projects/survival-shooter-tutorial
[2] www.thewalkingdead.com
[3] www.en.wikipedia.org/wiki/Resident_Evil
[4] www.en.wikipedia.org/wiki/Left_4_Dead

the game parameters - the speed, health and spawn time of the enemies. The different versions were designed taking into account existing studies about flow during gameplay [13], [15], as well as informal feedback from users.

Each version was fine-tuned to induce in the player a specific mental state, such as anxiety, boredom, engagement and frustration. The choice of which mental states to induce was based on the works of Csikszentmihalyi [8], Gilleade and Dix [25], and Poels *et al.* [26]. Players are able to play a sandbox version of the game and develop their initial skills. According to the flow model in Figure 1, this takes away the need to create a game version focusing on apathy, since participants will have developed their skills for the task.

The versions and their variations are presented in Table I. In addition to the game parameters values of enemies' speed, health and spawn time stats, we also alter other variables in versions B and D. In version B, the player's speed is reduced, the pick-ups' spawn time is enlarged, the ammunition spawn probability is increased and the health pick-up probability is reduced. Such settings make the game less engaging, since the player cannot run fast, which is expected to induce boredom in the player, and avoid apathy, since the player has more ammunition to fight back. In version D, the player's shooting probability is reduced. The variation makes the game too hard, which is expected to induce frustration in the player.

TABLE I: Version name, the mental state it is intended to induce and the general characteristics of each game version.

Version	Mental State	Parameters Settings
A	Anxiety	Low enemy health High enemy speed Low enemy spawn time
B	Boredom	High enemy health Low enemy speed High enemy spawn time Low player speed High pick-up spawn time High ammunition pick-up probability Low health pick-up probability
C	Engagement	Medium enemy health Medium enemy speed Medium enemy spawn time
D	Frustration	Low enemy health High enemy speed Low enemy spawn time Low shooting probability

In order to validate the inducement versions, we performed a user testing phase. We collected users' feedback via a questionnaire to validate that our versions induced the target mental states on the player. Results suggest that, although not at a significant level, versions A, B and C induce anxiety, boredom and engagement, respectively. Moreover, version D induces frustration at a significant level. We also collected biofeedback while participants were playing the game. Since results show that we induced the target mental states, we can use physiological data collected to model our classifier.

III-B. Classification Framework

The main function of the classification framework is to detect the mental state of the user. The detection of the mental state is carried out by a classifier. As mentioned in the previous section, all data obtained from performing the previous tests to the users is the input for the classifier. We model and train the classifier with these data so that the final solution can distinguish between the chosen mental states. This section addresses the validation of the physiological measures to be used when distinguishing the mental state of the user and the modeling of the classifier. We also present the final architecture of this main component.

III-B1 Choosing Physiological Measures

We collected physiological data from thirty users with a Plugged Kit[5]. Before playing the inducement versions, we recorded data for the baseline values and players had access to the sandbox. Each participant played the different versions in a random order for ten minutes each.

Before analyzing data, we processed the physiological measures. Bitalino reads the analog signals from the participants and transforms them into digital signals. These signals are not meaningful *per se* so we have to convert them to the physiological measures presented in section II-B. Thus, we converted them to Electrocardiography (ECG), Electrodermal Activity (EDA) and EEG values. After that we used the BioSPPy toolbox[6] to compute the instantaneous HR and GSR, and EEG alpha, beta and theta bands.

We performed Analysis of Variance (ANOVA) tests to check which physiological measures allow us to distinguish the different mental states. Alpha and theta bands were not significant in the distinction between the version the player was playing. Yet, the beta band allows to distinguish version C from versions B and D. We were not able to have a brainwave band that could significantly distinguish between all combinations of two versions. Alpha and theta bands did not have values that were significantly different between versions, yet the beta band proved to have significant differences. We can conclude that this is the best band of brainwaves to use when we are trying to distinguish the mental state of the player.

Regarding HR and GSR, none could effectively differentiate between the mental states. Moreover, GSR produced too many missing values, therefore it was discarded. Since we did not aim to create a game which could only adapt to the player based on one physiological feature, we decided to further analyze the values of the alpha and theta bands and the HR. We decided to follow-up with Wilcoxon tests to see how significant the differences were in these three signals between the versions. Results showed that the measure that better complements the beta band is the HR, thus the classifier uses data from the beta band and the HR to output a representation of the mental state of the player.

---

[5]www.bitalino.com/en/plugged-kit
[6]www.github.com/PIA-Group/BioSPPy

III-B2 Choosing Features

In order to select the frequency with which the user's mental state was to be assessed, we conducted an informal inquiry of several test subjects and empirically adjusted the sampling time to twenty seconds.

As we mentioned before, frustration emerges at the limit of anxiety. Therefore, and taking into account that it would simplify the problem, we decided to merge the anxiety and frustration mental states. This way we have to address only three mental states (anxiety, boredom and flow) and the transition between these three can be easily set by adjusting the difficulty based on how the classifier reads the mental state. Also, we used the difference between the values measured while the participant was playing the game and the mean of the control record. These allows us to create a classifier that addresses every user instead of a specific one, since it uses the differences between mental states rather than the real values.

Two of the features we decided to use are the mean and its variance. We chose them due to the fact that their combination can represent the shape of a distribution. The other three features are based on a Fast Fourier Transform (FFT). Each of them is composed by the sum of all real values that appear inside a certain frequency range. The intervals were chosen based on the shape of the real values of the FFTs and that artifacts would not be in the center of the graph. Thus, the number of dimensions of our data is ten.

III-B3 Choosing a Classifier

The goal of the classifier is to take the biofeedback and output a representation of the mental state of the user based on it. We adopted a methodology similar to [20], [27]–[29] and tested four classification algorithms: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Multilayer Perceptron (MLP). We divided the data for the training and testing set as 80% and 20%, respectively. Each of the four classifiers was modeled using a 10-fold cross-validation grid search method with a set of hyper-parameters with the scikit-learn toolkit[7]. Tables II and III show all the relevant values in which we based our choice.

TABLE II: Comparison of the classifiers with average test set results for accuracy, precision, recall and f1-score.

	Accuracy	Precision	Recall	F1-score
DT	0.72	0.75	0.75	0.75
RF	0.85	0.86	0.85	0.85
SVM	0.83	0.83	0.83	0.83
MLP	**0.87**	**0.87**	**0.87**	**0.87**

We used accuracy, precision, recall, f1-score and support as metrics to compare the classifiers. Accuracy is complemented by precision and recall because they allow us to have a more specific view of the performance of each classifier with each state. Moreover, precision and recall are adequate for datasets with a large skew in the class distribution [30], which is the case of the boredom dataset.

TABLE III: Comparison of the classifiers with average test set results for support to each condition.

	Sup. Anxiety	Sup. Flow	Sup. Boredom
DT	298	294	**54**
RF	**336**	285	25
SVM	302	**304**	40
MLP	310	292	44

The MLP showed higher accuracy, precision, recall and f1-score compared to the other three classifiers, although the differences against the RF and SVM are not significant. One reason why it happened may be related to the fact that we only use ten features and the classifiers reached a maximum value for all metrics with our dataset. Regarding support, RF has the higher value for anxiety, SVM for flow and DT for boredom, yet we have no preference to assess one mental state over the others. Although MLP does not have a higher support for any mental state, no other classifier dominates these three last measures as the MLP dominates the first four. This leads us to conclude that the best classifier for our work is the MLP.

III-B4 Architecture

The final architecture of the classification framework is composed by the Bitalino, the BioSPPy, and the feature extraction and classification algorithms. The flow of information in this framework is straight forward. First, users produce analog signals which are measured with the Bitalino. These signals are then digitized by OpenSignals (r)evolution and fed to the BioSPPy which filters them, performs R-peak detection for the computation of the instantaneous HR and EEG band division for the beta waves. After that, the feature extraction algorithm takes both the beta band and the HR and extracts the relevant features, which are fed to the classifier. The process occurs for time intervals of twenty seconds. At each twenty seconds, a bash script runs the BioSPPy toolkit and processes the text file from OpenSignals (r)evolution. BioSPPy outputs two text files: one has the instantaneous HR and the other the values for the beta band. Next, the bash script calls a Python scripts which extract the features from those files, load the classifier and write on a file the classified mental state.

III-C. Controller Framework

We implemented a state machine in our adaptable game which allows the game engine to know the current state and, when it receives the up-to-date mental state of the user, a transition to the next state depending on it. This section presents how the state machine was developed and how the game conducts the adaptation of the game parameters and environment settings.

III-C1 State Machine

We opted to create a game that has three different levels. The mental state of the player is the trigger that allows the game to travel between the different levels while they are playing. The transitions the state machine (see Figure 3) receives as input are the output of the classification algorithm. At each specific time interval, the game checks which mental state the classifier chose and feeds it to the state machine. The

state machine processes the mental state, transits to the next state accordingly to it and the game adapts itself based on the final state. The difficulty and number of engagement elements increase as the level number increases as well. It is similar to the one developed by Rani *et al.* [29]. As we mentioned in Section I, we would compare this state machine against one that would adapt to the player's performance. We created a second version of the game with a state machine that adapts to the performance.
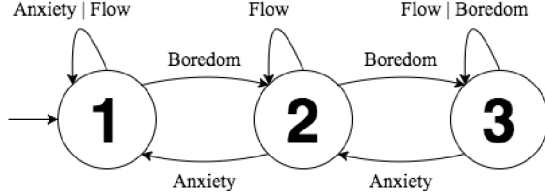

Fig. 3: Three states machine diagram for mental state-based gameplay adaptation. The player's mental state is classified as one of the following states: Anxiety, Boredom and Flow. When the game starts, the player starts at level 1.

Afterwards, we created the guidelines for the adaptation. The design criteria to keep the user in a flow state are focused on two different fronts: environmental settings to adapt engagement and enemies stats alteration to adapt the difficulty.

III-C2 Environmental Settings

As mentioned before, we needed to create elements in the game that would lead the player to a higher engagement and, consequently, to flow. The design guidelines for the environmental settings are the following: (1) In levels 2 and 3, when enemies, die, they explode giving Area of Effect (AOE) damage to nearby enemies, (2) In level 3, barrels spawn randomly on the environment and give AOE damage to nearby enemies when they explode, (3) Color of explosions depends on the flow of the player, and (4) Fitting sensory effects (explosions, sounds, textures).

With these guidelines, we change the way the player interacts with the enemies thereby creating engagement. The objective is to not prolong the task of killing an endless wave of zombies, since that can lead to boredom. This way, the player can create new strategies based on the level. The differences in the color of explosions is a passive way of showing the player that the level changed.

III-C3 Difficulty Settings

Regarding the difficulty adaptation, as higher the level, the difficulty is also harder. We decided that enemies gain speed, health and a shorter spawn rate as the player goes up in levels, gradually increasing combat difficulty. The guidelines presented in the previous section allow the player to fight harder enemies, since killing one or shooting at a barrel allows the player to deal damage to any number of surrounding enemies. This balance is responsible for creating a higher engagement although the enemies are harder to kill.

IV. Validation

IV-A. Method

As mentioned in Section I, we need to verify that the prototype that adapts to a representation of the mental state of the player provides a better gaming experience compared to one that adapts to their score. In order to do so, we use the self-perceived flow and in-game scores as metrics. First, we require that the self-perceived flow is higher when the user is playing a game that adapts to their mental state comparing to playing a game which gameplay adapts to their performance. We can then create **Hypothesis 1** (H1): *users have a higher self-perceived flow playing the mental state-based version compared to the performance-based.*

Second, we address if players have a higher performance when they play a game that adapts to their mental state compared to playing a game which gameplay adapts to their performance. We decided to compare this performance by the sum of all scores players would have by the end of each time they played. When an enemy was killed it provided a certain amount of points to the player. If it was a Zumbunny or a Zombear, the player would win 10 points; and 50 points if it was an Hellephant. We summed all the scores each player obtain while playing each version and created **Hypothesis 2** (H2): *users have higher scores playing the mental state-based version compared to the performance-based.*

In this section, we describe the population sample used in our study, material we used in the tests and the procedure used in our second testing phase.

IV-A1 Participants

We recruited subjects through standard procedures including direct contact and through word of mouth. Subjects included anyone interested in participating with at least 18 years old. Each participant was asked to sign a consent form. There were no potential risks and no anticipated benefits to individual participants.

We conducted a total of 21 tests. All tests occurred between 08:00h and 20:00h. The participants (16 males, 5 females) were ranged in age from 19 to 27 ($M = 22.43, SD = 1.91$). From all the participants, eleven had already been present in the first testing phase. Statistical tests reported no difference between data from player that played the game before and players that did not do it.

Only one participant reported no video game-playing time. The other participants play at least once a day (28.57%), at least once week (42.86%) or at least once a month (23.81%). Fifteen participants frequently play FPS. Again, statistical tests reported no difference between data from players with different FPS gaming periodicity.

IV-A2 User Evaluation

This testing phase had a consent form and an adapted version of the Game Engagement Questionnaire (GEQ) [31]. We used five items from the adapted GEQ. Four address flow and one presence. We use them to see which version led to a higher self-perceived flow.

IV-A3 Apparatus

Two computers were used for this testing phase. Data was recorded at 100 Hz with OpenSignals (r)evolution Mac OS X (v.2017)[8] software. Both computers were capable of processing and running the programs without any delay. Players interacted with the game through mousepad and headphones. We also carefully chose a set of images, most of them from the International Affective Picture System (IAPS) [32] to lead to a relaxing state, as done in [33], [34].

We used a BITalino[9] to record the Blood Volume Pulse (BVP) with a photoplethysmography sensor and the electricity on the forehead with three electrodes. The first allows us to know the instantaneous HR and the second the beta band. While they were playing, participants had an earplug attached to one of their ears and two electrodes on their forehead (positions FP1 and FP2 in the 10–20 system [35]), one electrode on the left side of their neck (its function is to serve as "ground" for the difference in both hemispheres). Users were seated on a chair with the computer in front of them on top of a table, as seen in Figure 4. Other materials used for the experiences were a headband, neurodiagnostic electrode paste, alcohol and cloth.

Fig. 4: Participant playing the game.

IV-A4 Procedure

The tests were conducted in a laboratory inside campus Alameda, in Instituto Superior Técnico. The assistant started by explaining to the users the purpose of the study, what they would be doing and that they should not move their head while they were playing in order to prevent detaching the electrodes. After it, we asked the users to fill a consent form and a form regarding their demographic data and gaming experience. Then, we placed the physiological sensors on the user. We recorded the baseline values for five minutes while players looked to the chosen set of images.

After that, users played a sandbox version of the game to try the sensitivity and in-game interactions. The assistant allowed users to play as much time as they wanted so that they could develop the minimum skills to play the game in the different versions. Even users that had participated in the first testing phase had to play to ensure that participants knew and remembered all in-game interactions. Each tester

[8]www.bitalino.com/en/software
[9]www.bitalino.com/en/hardware

played each version for ten minutes and the playing order was randomized between players, meaning that players would either play first the mental state-based adaptable game and after it the performance-based adaptable game or vice-versa.

The procedure used for each version the participant played was the following:
1) The assistant verified if every sensor was correctly placed;
2) The assistant asked the tester if they was comfortable and ready to play;
3) The assistant started the game for the tester;
4) If the player died, the game automatically restarted;
5) The player played for ten minutes and after it the game automatically closed;
6) The assistant asked the tester to fill the form addressing the version the latter played;
7) The player rested for three minutes looking at a picture in order to return to a neutral mental state; and
8) After three minutes, the assistant repeated this procedure to the next version, if there was any other version to play.

The order of the adapted GEQ items was randomized between the two versions. We kept the gaming closing itself so that we would not break the participant's immersion in the game. After the tester played the two versions, the assistant removed the sensors from the tester and cleaned them with alcohol to remove the electrode gel. Free comments were also invited. Testers received a compensation based in candies. We created a contest to see which player achieved the highest score across all gaming sessions. The winner received a gift card worth 20EUR.

IV-B. Results

We used data obtained from the Adapted GEQ to address hypothesis **H1**. We conducted a paired-samples t-test to compare the self-perceived flow using values between the mental state-based and the performance-based versions. There was a significant difference in the self-perceived flow for the performance-based ($M = 6.46, SD = 1.29$) and mental state-based ($M = 6.05, SD = 1.34$) conditions; $t(20) = 2.215, p = 0.039$. These results show that the version the participant played has an effect on their self-perceived flow, more specifically that the performance-based prototype leads to a greater self-perceived flow compared to the mental state-based one. Therefore, we refute our hypothesis. Although both conditions have a similar Interquartile Range (IQR) (therefore, a similar variation), we can observe in Figure 5 that the median of the performance-based version has a slightly higher value compared to the mental state-based version.

Regarding **H2**, a paired-samples t-test was conducted to compare the scores in the mental state-based and the performance-based versions. There was a significant difference in the scores for the performance-based ($M = 4820, SD = 1828.52$) and mental state-based ($M = 4326.67, SD = 1376.35$) conditions; $t(20) = 2.635, p = 0.016$. These results show that the version the participant played has an effect on their score, more specifically that users reach higher scores playing the performance-based prototype compared to the
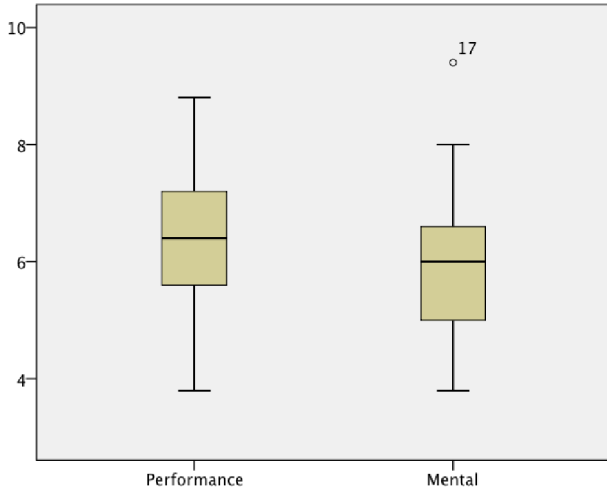
Fig. 5: Boxplots of normalized Flow Degree for performance and mental state-based versions.

mental state-based one. Therefore, we refute our hypothesis. Additionally, Figure 6 depicts the distributions and we see that the performance-based version has a higher median compared to the mental state-based version. Yet, the mental state-based condition has a much smaller IQR which suggests that the variation was smaller compared to the one of the performance-based condition.
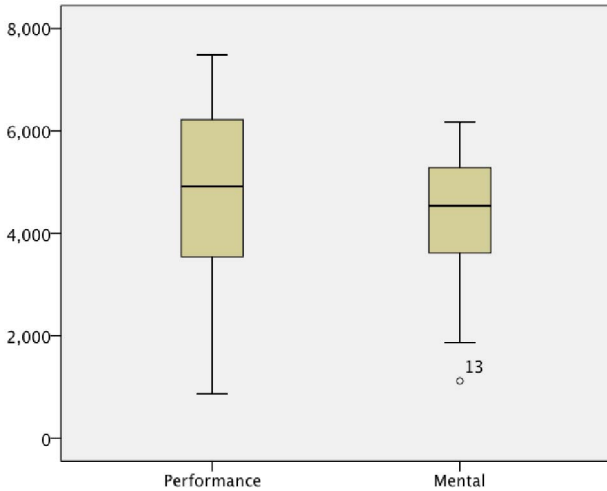

Fig. 6: Boxplots of scores for performance and mental state-based versions.

V. DISCUSSION

Although there are empirical evidences that users would have a higher flow state and scores playing a game that adapted itself to their mental state, after our testing phase, we found out that they have a higher flow state and scores playing a game that adapts to their performance. A deeper statistical analysis confirmed that players have a higher flow state and scores with a performance-based adaptable game, so in fact players do not

have a higher flow state and scores with a mental state-based adaptable game. Even if not supported by the data, we believe that our hypothesis is valid, since it is based on the Flow Theory [8] which has been well studied in recent decades.

One reason why we could not fully validate our hypothesis may be related to the data we used to model our classifier. When we collected it, the number of participants was small. A higher number of participants would allow to generalize for the whole population and take conclusions with a stronger impact. In future studies, a bigger sample size must be acquired. Also, the number of female participants was too small compared to the number of male participants, so future samples should consider increase female presence as well. These problems also appeared in the second testing phase, but with a smaller sample of participants and another disparate ratio of male and female gamers.

Moreover, the game periodicity of the players in both testing phases was not proportionally represented across all the possible ones. Only in the first testing phase we found that the game periodicity influenced the anxiety in version C, yet, as it was only one case in sixteen possible cases, different gaming periodicity regarding FPS was assumed to be minimal.

Another limitation was the OpenSignals (r)evolution software. Since it did not allow a real-time recording, the assistant had to keep on restarting the recording after each twenty seconds. This was not practical and led to considerable small delays on the update of the file being processed by the classification framework. Nonetheless, these delays were rare or no longer than forty seconds, thus we consider them irrelevant.

Our findings provide additional evidence for inducing and investigating different mental states achieved while playing computer games. In particular, game developers may use the game design options to induce anxiety, boredom, flow and frustration or to maintain an optimal challenge based on the players mental state or performance. These mental states can be useful to provide a better gaming experience with task difficulty adjustment.

VI. CONCLUSIONS

We investigated whether flow may be relevant for gameplay adaptability and may offer a better gaming experience, since that mental state is associated with deep engagement. If so, flow-based adaptation can be a further step in serious games for health investigation by providing a framework that keeps patients more engaged in the recovery process.

Future work involves creating a larger data set with a higher number of participants of both genders. This way we can generalize for the whole population and model a classifier with higher accuracy. The number of participants that play FPS must be increased, since they are used to the game type and can provide a better feedback than players who only play mobile games, for example. Another approach is to complement our dataset with more representations of mental states of the user in order to provide a wider set of adaptability components and address more dimensions other than skills of the player and challenge of the game. Further development of

the MLP classifier involves using different features from the ones we chose and choosing different physiological signals. Regarding software we used in our study, there is future work using the OpenSignals (r)evolution application programming interface to create a real-time recording of the physiological signals, which may prove to be a more effective way to access the participants' physiological signals without the disruption of restarting the recording at each time interval.

Acknowledgment

References

[1] K. Durkin, "Videogames and young people with developmental disorders," *Review of General Psychology*, vol. 14, no. 2, p. 122, 2010.

[2] A. Krichevets, E. Sirotkina, I. Yevsevicheva, and L. Zeldin, "Computer games as a means of movement rehabilitation," *Disability and rehabilitation*, vol. 17, no. 2, pp. 100–105, 1995.

[3] E. Flores, G. Tobon, E. Cavallaro, F. I. Cavallaro, J. C. Perry, and T. Keller, "Improving patient motivation in game development for motor deficit rehabilitation," in *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*. ACM, 2008, pp. 381–384.

[4] D. J. Reinkensmeyer and S. J. Housman, "If i can't do it once, why do it a hundred times?: Connecting volition to movement success in a virtual environment motivates people to exercise the arm after stroke," in *Virtual Rehabilitation, 2007*. IEEE, 2007, pp. 44–48.

[5] M. K. Schlickum, L. Hedman, L. Enochsson, A. Kjellin, and L. Felländer-Tsai, "Systematic video game training in surgical novices improves performance in virtual reality endoscopic surgical simulators: a prospective randomized study," *World journal of surgery*, vol. 33, no. 11, p. 2360, 2009.

[6] S. Göbel, L. Salvatore, and R. Konrad, "Storytec: A digital storytelling platform for the authoring and experiencing of interactive and non-linear stories," in *Automated solutions for Cross Media Content and Multi-channel Distribution, 2008. AXMEDIS'08. International Conference on*. Ieee, 2008, pp. 103–110.

[7] J. G. Marin, K. F. Navarro, and E. Lawrence, "Serious games to improve the physical health of the elderly: A categorization scheme," in *International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services. Barcelona, Spain*, 2011.

[8] M. Csikszentmihalyi, *Beyond boredom and anxiety*. Jossey-Bass, 2000.

[9] S. Arnab, T. Lim, M. B. Carvalho, F. Bellotti, S. De Freitas, S. Louchart, N. Suttie, R. Berta, and A. De Gloria, "Mapping learning and game mechanics for serious games analysis," *British Journal of Educational Technology*, vol. 46, no. 2, pp. 391–411, 2015.

[10] M. Csikszentmihalyi, *Finding Flow: The Psychology of Engagement with Everyday Life*. Basic Books, 1997.

[11] ——, *Flow: The Psychology of Optimal Experience*. NY Harper & Row, 1990.

[12] J. Nakamura and M. Csikszentmihalyi, "The concept of flow," in *Flow and the foundations of positive psychology*. Springer, 2014, pp. 239–263.

[13] L. E. Nacke and C. A. Lindley, "Flow and Immersion in First-person Shooters: Measuring the Player's Gameplay Experience," *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, pp. 81–88, 2008.

[14] G. D. Ellis, J. E. Voelkl, and C. Morris, "Measurement and analysis issues with explanation of variance in daily experience using the flow model," *Journal of leisure research*, vol. 26, no. 4, p. 337, 1994.

[15] L. Harmat, Ö. de Manzano, T. Theorell, L. Högman, H. Fischer, and F. Ullén, "Physiological correlates of the flow experience during computer game playing," *International Journal of Psychophysiology*, vol. 97, no. 1, pp. 1–7, 2015.

[16] Y. Bian, C. Yang, F. Gao, H. Li, S. Zhou, H. Li, X. Sun, and X. Meng, "A framework for physiological indicators of flow in VR games: construction and preliminary evaluation," *Personal and Ubiquitous Computing*, pp. 1–12, 2016.

[17] L. E. Nacke, S. Stellmach, and C. A. Lindley, "Electroencephalographic assessment of player experience: A pilot study in affective ludology," *Simulation & Gaming*, 2010.

[18] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 6, pp. 1052–1063, 2011.

[19] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.

[20] R. Berta, F. Bellotti, A. De Gloria, D. Pranantha, and C. Schatten, "Electroencephalogram and physiological signal analysis for assessing flow in games," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 5, no. 2, pp. 164–175, 2013.

[21] A. McMahan, "Immersion, engagement and presence," *The video game theory reader*, vol. 67, p. 86, 2003.

[22] M. Grimshaw, *The acoustic ecology of the first-person shooter*. VDM Verlag, 2008.

[23] J. Takatalo, T. Kawai, J. Kaistinen, G. Nyman, and J. Häkkinen, "User Experience in 3D Stereoscopic Games," *Media Psychology*, vol. 14, no. 4, pp. 387–414, 2011.

[24] B. Cowley, D. Charles, M. Black, and R. Hickey, "Toward an Understanding of Flow in Video Games," *Computers in Entertainment*, vol. 6, no. 2, pp. 20:1–20:27, 2008.

[25] K. M. Gilleade and A. Dix, "Using frustration in the design of adaptive videogames," *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology - ACE '04*, pp. 228–232, 2004.

[26] K. Poels, Y. De Kort, and W. Ijsselsteijn, "It is always a lot of fun!: exploring dimensions of digital game experience using focus group methodology," in *Proceedings of the 2007 conference on Future Play*. ACM, 2007, pp. 83–89.

[27] T. I. Laine, K. Bauer, J. W. Lanning, C. A. Russell, and G. F. Wilson, "Selection of input features across subjects for classifying crewmember workload using artificial neural networks," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 32, no. 6, pp. 691–704, 2002.

[28] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R. E. John, L.-O. Wahlund, Y. Dodge, and T. Dierks, "Application and comparison of classification algorithms for recognition of alzheimer's disease in electrical brain activity (eeg)," *Journal of neuroscience methods*, vol. 161, no. 2, pp. 342–350, 2007.

[29] P. Rani, N. Sarkar, and C. Liu, "Maintaining Optimal Challenge in Computer Games through Real-Time Physiological Feedback Mechanical Engineering," *Task-Specific Information Processing in Operational and Virtual Environments: Foundations of Augmented Cognition*, pp. 184–192, 2006.

[30] R. Bunescu, R. Ge, R. J. Kate, E. M. Marcotte, R. J. Mooney, A. K. Ramani, and Y. W. Wong, "Comparative experiments on learning information extractors for proteins and their interactions," *Artificial intelligence in medicine*, vol. 33, no. 2, pp. 139–155, 2005.

[31] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. Mcbroom, K. M. Burkhart, and J. N. Pidruzny, "The development of the Game Engagement Questionnaire : A measure of engagement in video game-playing," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 624–634, 2009.

[32] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Affective ratings of pictures and instruction manual," *Technical report A-8*, 2008.

[33] P. J. Lang, "The emotion probe: Studies of motivation and attention," *American psychologist*, vol. 50, no. 5, p. 372, 1995.

[34] P. Lang and M. M. Bradley, "The international affective picture system (iaps) in the study of emotion and attention," *Handbook of emotion elicitation and assessment*, vol. 29, 2007.

[35] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution eeg and erp measurements," *Clinical neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.