**Understanding the user gameplay experience through two different game controllers**

Daniela Trevisan¹, Guilherme Gonçalves¹, Érica Mourão¹

¹Fluminense Federal University, Computing Institute. Rua Passos da Pátria 156, Niterói, Rio de Janeiro, Brazil.

{daniela,galves,ericamourao}@ic.uff.br

**Abstract.** Emerging technologies offer exciting new ways of interacting with digital games to create fantastic play experiences. Evaluating entertainment technology is challenging because success isn’t defined in terms of productivity and performance, but in terms of enjoyment and gameplay experience. In this direction, the success criteria here has been evaluated in an exploratory user experience and usability study that used a within non-players design investigating a commercial joystick as input device compared to a novel adaptive touch-based controller. This new adaptive controller is able to dynamically change buttons size and position according to the user behavior. The evaluation methodology is based on the capture and analysis of objective and subjective user´s data. Results of this exploratory study indicate that while the general user experience and usability were similarly perceived in both controllers the physiological measures indicate that the user's emotions were greater with the adaptive controller. The findings are subsequently discussed and implications of using the adaptive touch-based controller as input for digital games are discussed..

**Keywords:** User Experience based approaches; Evaluation Methods / Usability Evaluation; Human Factors and HCI; Adaptive Interfaces.

# **5. Results**

For ease of presentation and hence understanding, we present the experiment results in the following subsections. Besides, we applied a Wilcoxon signed ­rank test, a non­parametric statistical hypothesis test with significance level of 0.05 and the two-tailed hypothesis defined, returning the p-value, which indicates if the difference between the results achieved for both groups is significant. The decision of using a non­parametric test was motivated by the unknown distribution of the test results, since a parametric test requires a previous knowledge about the data distribution, which must be normal. For our groups' sizes, a p-value smaller than 0.05 represents a significant difference. Furthermore, for correlation results we applied a Pearson linear correlation measure expecting to find some correlations between our variables.

## **5.1. Attractiveness**

For the hypotheses H1 (The user experience on adaptive control is greater than the traditional one, based on Attrakdiff questionnaire results), the Figure 2 shows the obtained results from Attrakdiff in its four dimensions, with each the respectively means and standard deviations. The Pragmatic Quality dimension (PQ) is perceived as greater in traditional controller (M = 3.89, SD = 0.72) than the adaptive one (M = 3.99, SD = 1.21). Although, the Wilcoxon test for the hypothesis H1 returned a non statistically significant result, since p-value = 0.661. The Hedonic quality - identity (HQI) is also greater in traditional controller (M = 4.49, SD = 1.47) than the adaptive one (M = 4.40, SD = 1.03). Although, the Wilcoxon test for the hypothesis H1 returned a non statistically significant result, since p-value = 0,556. The Hedonic quality - stimulation (HQS) is also greater in traditional controller (M = 4.91, SD = 1.72) than the adaptive one (M = 3.79, SD = 0.81). Although, the Wilcoxon test for the hypothesis H1 returned a non statistically significant result, since p-value = 0,989. The Attractiveness dimension (ATT) as well as the others, is greater in traditional controller (M = 3.93, SD = 0.71) than the adaptive one (M = 3.86; SD = 0.91). Although, the Wilcoxon test for the hypothesis H1 returned a non statistically significant result, since p-value = 0,743. As seen in Figure 5a, all results found for hypothesis H1 returned a non statistically significant result. The Figure 5b show the mean in scale Likert -3 to 3.

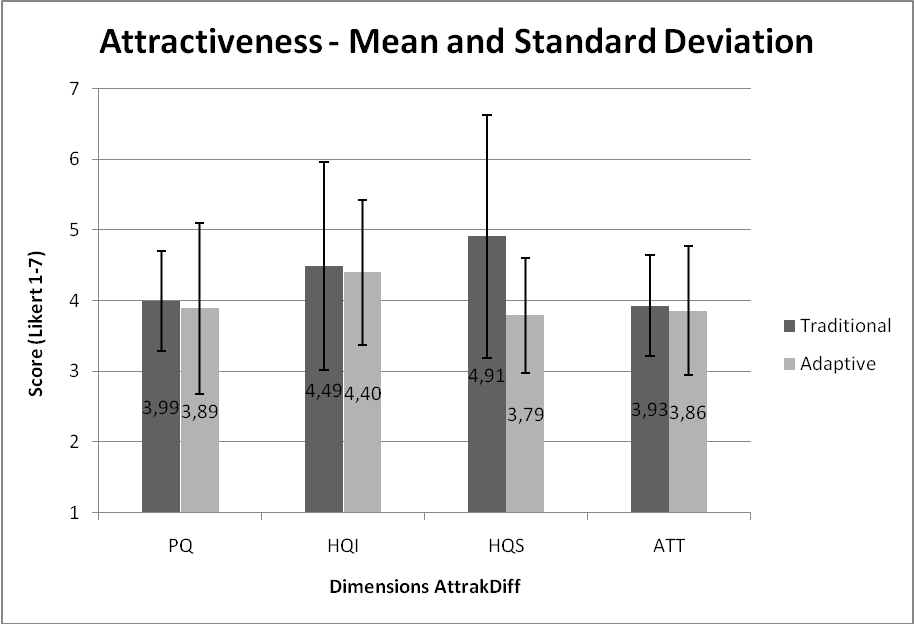


Figure 5a:AttrakDiff Questionnaire Results – Mean and Standard Deviation

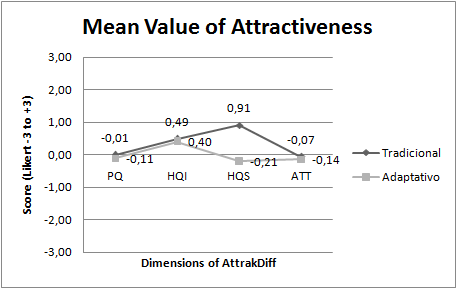


Figure 5b:AttrakDiff Questionnaire Results – Mean in Likert -3 to 3

## **5.2. User´s Preference**

For the hypothesis H2 (Overall, the user preference is greater on adaptive controller rather than the traditional one.), the results show that the UX during the experiment is perceived as lower in adaptive controller than traditional controller. Comparing the mean score from SUS in adaptive controller (64.8) and the traditional controller (74.0) with 5% of margin of error, we perceived that the score in adaptive controller is below the desirable mean (68.0) . When we paired the participants SUS scores, the Wilcoxon results show a non significant p-value of 0.853, rejecting the hypothesis H2.

## **5.3. Performance**

For our hypothesis H3 (In general, performance is similar between adaptive and traditional control.) we have the results obtained according to performance metrics (Stage duration in seconds, quantity of remaining player’s lifes at the end of stage and number of player’s deaths). For the metric stage duration in seconds (stage 1) was observed a lesser value in traditional controller (M = 145,4) than the adaptive one (M = 158,7). In Stage 2 the traditional controller is greater (M = 268,3) than the adaptive (M = 243,9). Although, the Wilcoxon test obtained on this metric for both stages, respectively, resulted on a p-value of 0.375 and 0.444. This results shows a significant difference between the results for this analysis. Figure 6 illustrates the duration averages (in seconds) of each game stage and their respective standard deviations.

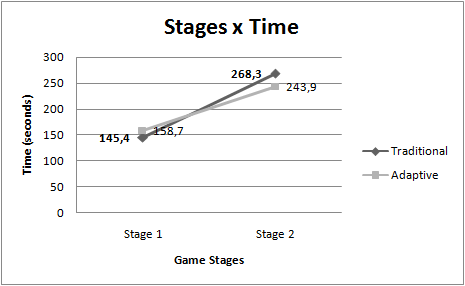


Figure 6: Mean Time (seconds) for each game stages.

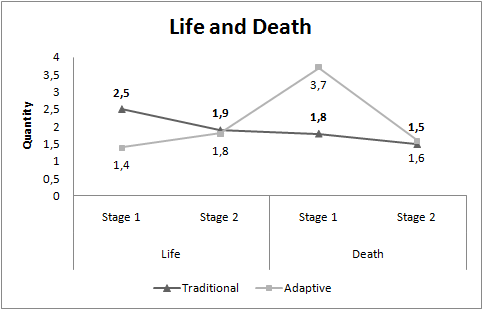


Figure 7: Mean Time (seconds) for each game stages.

Quantity of remaining life at the end of stage 1 is perceived as greater in traditional controller (M = 2.5) than the adaptive (M = 1.4) and at stage 2 also is greater in traditional controller (M = 1.9) than the adaptive (M = 1.8). However the number of player deaths is greater in adaptive controller in stage one (M = 3.7) than traditional (M = 1.8) and in stage second (M = 1.6) than (M = 1.5) respectively. Regarding the metric (quantity of remaining life at the end of each stage), we found a significant difference only for the first one, obtaining a p-value of 0.020, while in the second one, we had a non-significant p-value of 0.821. For the metric (number of player deaths), no significant difference was evidenced in both stages, since the p-values where 0.188 and 0.930. In this scenario, the only statistically significant result indicates a performance improvement in stage 1 from the game using the adaptive control. The Figure 7 show us the results.

## **5.4. Emotion analysis**

For the hypothesis H4 (User’s emotions are similar in both controllers during the experiment.), we compared the emotions (Engagement, excitation, interest, relaxation, stress and focus) results between both controllers. Our findings show that all emotions are greater in adaptive controller. The engagement in adaptive controller scored higher levels (M = 57.5, SD = 3.98) than the traditional controller (M = 57.4, SD = 4.30). The excitement also got greater levels in adaptive controller (M = 20.3, SD = 7.10) than the traditional one (M = 18.1, SD = 6.62). The interest, in the same way was greater in adaptive controller (M = 56.1, SD = 3.73) than the traditional one (M = 55.7, SD = 3.06). The relaxation as well as the others was greater in adaptive controller (M = 31.2, SD = 3.22) than traditional one (M = 31.1, SD = 2.51). The stress was perceived as greater in adaptive controller (M = 48.0, SD = 12.96) than traditional one (M = 46.7, SD = 11.79). Finally, the focus was also greater in adaptive controller (M = 34.8 SD = 9.76) than traditional controller (M= 32.6, SD = 8.26). Overall emotions are greater in adaptive controller. We verify if exists significant difference. The Figure 8 shows the means and standard deviations from each emotions recorded from Emotiv epoc + device. After results from Wilcoxon show a significant p-value (< 0.05) only in excitement (0.0141) and focus (0.0245) emotions.

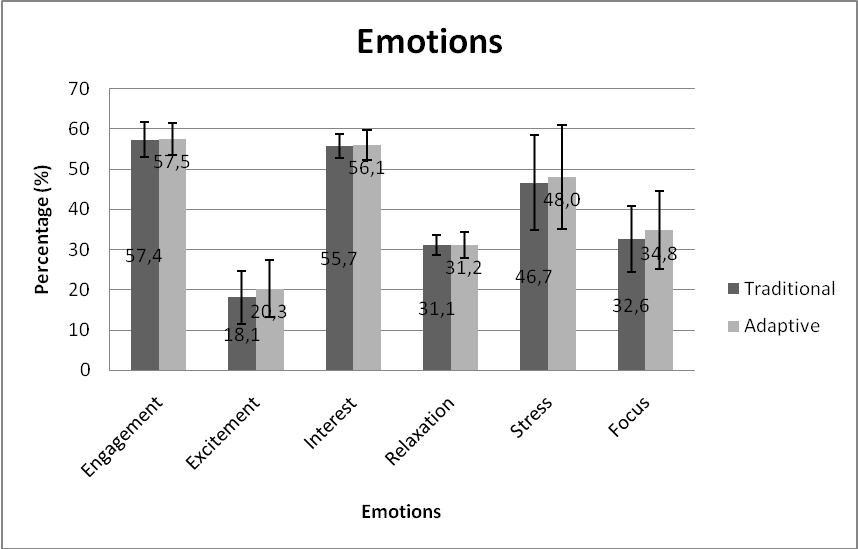


Figure 8: Emotions - Mean and Standard Deviation.

We observed, according to table 1, that the excitement and focus emotions are those that obtained p-value below 0.05, which indicates that there is significant difference. For the other emotions in this study, i.e. engagement, interest, relaxation and stress, no significant difference is evidenced, which means that these emotions are similar in both controllers.

|  |  |
| --- | --- |
| **Emotions** | **p-value** |
| Engagement | 0,8539 |
| Excitement | **0,0141** |
| Interest | 0,5245 |
| Relaxation | 0,8539 |
| Stress | 0,3997 |
| Focus | **0,0245** |

Table 1: Wilcoxon p-values for each emotion.

## **5.5. Correlation between emotions and performance**

For the hypothesis H5 (User’s emotions on adaptive control do not affect the performance), we tested the correlation between performance and emotion using the Pearson linear correlation, which a strong correlation is defined as a p-value lower than 0.05. A significative difference in number of player’s deaths metric was found in stage 1. The Pearson value from excitement and interest emotions and this performance metric shows a moderate correlation (as seen in Table 2).

|  |  |  |  |
| --- | --- | --- | --- |
| **Emotion versus Number of player’s deaths** | | | |
| Emotions | **Controller** | **P-value < 0,05** | **Pearson Correlation**  **0,5 <= r < 0,8** |
| Excitement | Traditional | 0,0065 | 0,73 |
| Adaptive | 0,0161 | 0,78 |
| Interest | Traditional | 0,0242 | 0,69 |
| Adaptive | 0,0670 | 0,59 |

Table 2. Emotions versus Number of player’s deaths in the Game (Only statistically significant (p < 0.05) results are reported.)

For stage 2, the emotions "excitement" and "interest" have significant differences regarding the metrics "stage duration in seconds" in the use of both controllers (see table 3). Therefore, the Pearson coefficient evidenced a moderate positive value for "interest" and a strong positive value for "excitement", which indicates a correlation for each emotion.

|  |  |  |  |
| --- | --- | --- | --- |
| **Emotion versus Stage duration in seconds** | | | |
| **Emotions** | **Controller** | **P-value < 0,05** | **Pearson Correlation**  **0,5 <= r < 0,8 moderate 0,8 < = r < 1 stronger** |
| Excitement | Traditional | 0,0016 | 0,85 |
| Adaptive | 0,0044 | 0,81 |
| Interest | Traditional | 0,0466 | 0,64 |
| Adaptive | 0,0322 | 0,67 |

Table 3. Emotions versus Stage duration in seconds

Brain waves identified in Brain Activity Map are: Delta, Theta, Alfa and Beta for all sensor location on Emotiv Epoc headset. Delta waves (0.5-4 Hz) indicate, when active, deep sleep, rest and, conversely, agitation when suppressed. Theta (4-8Hz) are waves that indicate state of deep, dream meditation. Alpha waves (8-15Hz) indicate states of relaxed alertness, rest and meditation. And Beta waves (15-30HZ) indicate wakefulness, alertness, mental engagement, conscious information processing [19]. Some difference in the EEG might be explained by a mixed of emotions. An increase in stress might also fell some kind of excitation [20, 18].

The Figure 9 can be interpreted as follows: the red color in Theta wave is linked to vigil with forced attention state (concentration) to relaxation before the game. Alpha is linked to wakefulness with relaxation. Beta associated with vigil with state of attention.

The Figures 10 and 11 can be interpreted as: the red color in Alpha and Beta indicates an increase in frequency in these waves.

The results show that the Alpha wave had an increased frequency in both controls at the time of the player's death but the activity was the same in adaptive and traditional controller.

However in the Beta wave increase was greater in adaptive control than in traditional. Beta indicates that there was an increase in activity associated with intense mental concentration. The emotion excitement can be explained by the increase in Beta frequency. As the red coloration is lower in the traditional control for the Beta wave, we understand that the excitation was lower.

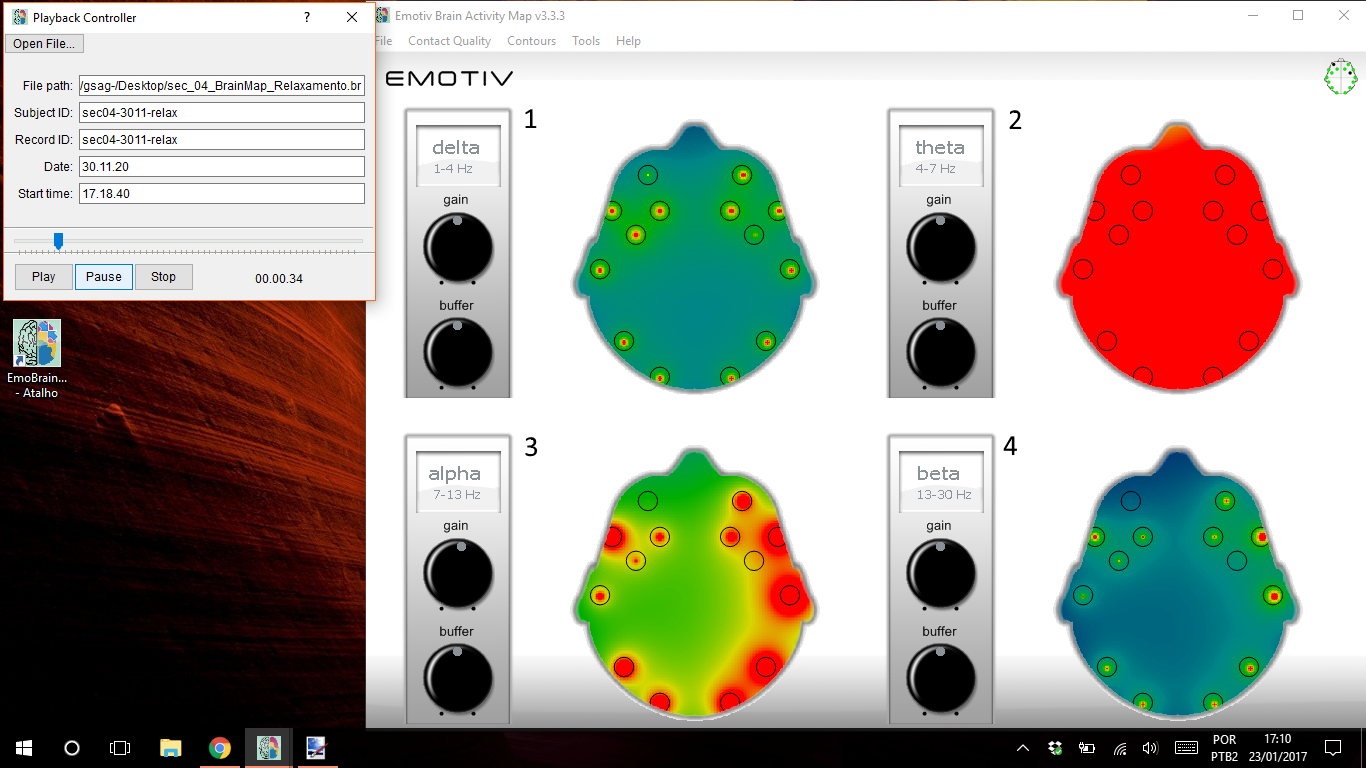


Figure 9 – Relaxation before the game (Neural baseline)



Figure 10 - Adaptive – Neural activity in the moment of player’s death.



Figure 11 - Traditional - Neural activity few seconds before player’s death.

## **5.6. Correlation between emotions and user experience**

For hypothesis H6 (When attractiveness raises, interest emotion levels increases), we take the results from AttrakDiff dimensions and the emotions. The correlation between the Pragmatic Quality dimension and the Interest emotion was strong (p-value = 0.0066) with value 0.7894 of moderate correlation. There were no significant differences between other emotions that indicated correlation. With this, we can conclude that there is no correlation between the the hedonic quality (HQI) and the interest emotion.

## 5.7. Correlation between attractiveness and user preference

For hypothesis H7 (User’s preference induce positively on attractiveness dimensions), further investigation of the interrelations between different indicators were computed using Pearson correlation coefficient. As a reference for usability and user experience attractiveness, respectively, the analysis included the items from the SUS questionnaire and the Attrakdiff questionnaire dimensions.

The SUS rating of the controllers showed a non significant correlation at p < 0.05 with all attractiveness dimensions. For the adaptive controller we obtained the following values (PQ -0.445, p = 0.197; HQI 0.182, p = 0.614; HQS -0.445 p = 0.197; ATT 0.449, p = 0.193) and for the traditional one we obtained the following values (PQ 0.283, p = 0.428; HQI 0.384, p = 0.273; HQS -0.254, p = 0.478; ATT 0.023, p = 0.949). In general, the results suggest that the attractiveness and user preference factors are weakly interrelated. It is important to investigate in more detail how the attractiveness influences the perception of the user preference, or if the two concepts are independent.

# **6.** Discussion

The study reported here is the first experiment to compare an innovative controller and a traditional one. At current stage we have to deal with the following limitations: (1) the few amount of performance variables from the game and controllers; (2) the small sample size (10 participants); (3) the emotion and performance analysis could be complemented with other physiological measures i.e Electromyography (EMG). We believe that future works with a greater sample size and more performance variables could bring some news insights. As well as, new physiological measures helped to try to better understand the results.

The results from Attrakdiff questionnaire in its four dimensions is perceived as greater in traditional controller than adaptive controller. However the user experience, in the context of attractiveness, did not show a significative difference between the adaptive controller and the traditional one in our hypothesis (H1). Which means that the user experience in both controllers are perceived as similar.

The results of the hypothesis (H2) indicate that the user preference is greater in traditional controller than adaptive controller but has no significant difference between the controls. This shows that the user’s preference is also similar in the adaptive controller and the traditional.

The user experience and the user preference may have been affected by the participants' profile: for the inexperienced players, both controls are a new experience. And the number of participants may have influenced the results statistically. We believe that greater game duration would bring more significant data.

We observed in the performance metric (stage duration in seconds) in hypothesis (H3) that the participants concluded each stage with different durations. The Wilcoxon test showed significant differences between controllers in stages durations. The first stage took more time with adaptive controller than traditional and the second stage of the game took less time with adaptive controller than traditional controller. This leads us to think that in stage one the participant took more time to adapt to adaptive control. With this in stage second the control already adapted to the touch.

The duration of second stage was greater than of first stage for the both adaptive and traditional controllers. This can be explained by the presence of many obstacles in the second stage of the game. To not die the player needs deflect of the obstacles used only directional buttons up and down.

Regarding the performance metrics (quantity of remaining player’s lives at the end of stage) and (number of player’s deaths) in hypothesis (H3) we observed the results.

The quantity of remaining player’s lives at the end of each stage was greater in traditional control than adaptive control. The significant different was perceived only in stage one. Maybe the player adapting to the touch of adaptive control in stage on. With this, more life would be left in the use of traditional control.

The number of player’s deaths was greater with the use of adaptive control than the use of traditional control in both stages. However, no significative difference was evidenced. In both controls, adaptive and traditional, the player died more times in stage one than second stage. We believe that the player paid more attention to the obstacles of the second stage game to die less.

In this scenario presented of the hypothesis (H3), the only statistically significant result indicates a performance improvement from stage one to stage second in the game using the adaptive control and we concluded that the performance is not similar in controls.

The work reported here show results that user’s emotions are not similar in both controllers during the experiment. The emotions of experiment (Engagement, Excitement, Interest, Relaxation, Stress, and Focus) are greater in adaptive controller than traditional controller but wilcoxon test show significant difference only in excitement and focus. We believe that the excitement was greater in adaptive control by being a different, new control in a smartphone handset. Already the emotion focus was greater on adaptive control because it did not have a tactile, physical button and overall, you would not need to look for control during the game. This way the player would remain focused on the game screen. Looking into the all emotions of the study we concluded to hypothesis (H4) that some user’s emotion are similar and other are not similar in both controllers.

The results show that the correlation analysis between emotions and performance exists. First a significative difference was found to emotions (excitement and interest). Pearson correlation value shows moderate correlation in emotions versus number of player's death in the stage one of the game. Also the emotions "excitement" and "interest" have significant differences regarding the metrics "stage duration in seconds" in the use of both controllers in second stage. Therefore, the Pearson coefficient evidenced a moderate positive value for "interest" and a strong positive value for "excitement". This results of hypothesis (H5) show that emotions excitement and interest affect the performance number of player's death and stage duration in seconds.

Results from AttrakDiff dimensions and the emotions show the moderate correlation between the Pragmatic Quality dimension and the “Interest” emotion. Others dimensions and emotions not have correlation for hypothesis (H6). This may indicate that the interest raises when the quality of an application raises and the degree of success that users achieve the goals.

Additional analysis between SUS rating and AttrakDiff dimensions in hypothesis (H7) shows in general that user preference and attractiveness do not correlate. The profile of the participant with little or no experience and the style of the game could have influenced the results. An in-depth study could bring more significant results.

Unfortunately, we did not found more significative aspects between the adaptive and traditional controllers that evidenced others conclusive answers to our research questions. We hope with our study that the contribution made may inspire other researchers to explore this field of science.

# 8. References

1. Bernhaupt, R. 2010. User experience evaluation in entertainment. Evaluating User Experience in Games. Springer London, pp. 3-7.
2. Hotho, S., McGregor, N. 2013. Changing the Rules of the Game: Economic, Management and Emerging Issues in the Computer Games Industry. Springer.
3. YoYo Games: GameMaker. http://www.yoyogames.com/gamemaker. Accessed 23/11/2016.
4. Torok, L., Pelegrino, M., Lessa, J., Trevisan, D.G., Vasconcelos, C.N., Clua, E. & Montenegro, A. 2015. Evaluating and customizing user interaction in an adaptive game controller. Design, User Experience, and Usability: Interactive Experience Design, Springer, pp. 315-326.
5. Malfatti, S. M., Dos Santos, F.F. & Dos Santos, S.R. 2011, Using mobile phones to control desktop multiplayer games, Proceedings - 2010 Brazilian Symposium on Games and Digital Entertainment, SBGames 2010, pp. 230.
6. Vajk, T., Coulton, P., Bamford, W., Edwards, R., 2007. Using a Mobile Phone as a “Wii-like” Controller for Playing Games on a Large Public Display, International Journal of Computer Games Technology 2008*,* e539078.
7. Torok, L., Pelegrino, M., Trevisan, D.G., Clua, E. & Montenegro, A. 2015, A mobile game controller adapted to the gameplay and user’s behavior using machine learning. Entertainment Computing - ICEC 2015, Springer, pp. 3-16.
8. Pelegrino, M., Torok, L., Trevisan, D. & Clua, E. 2014, Creating and Designing Customized and Dynamic Game Interfaces Using Smartphones and Touchscreen, Brazilian Symposium on Games and Digital Entertainment, SBGAMES, pp. 133.
9. CICTEC, Centro de Inteligência Competitiva para Parques Tecnológicos. Tendências de Inovações para a Indústria de Jogos Digitais. 2013. Disponível em: <http://pt.slideshare.net/cictec/inovaes-tecnolgicas-para-a-industria-de-jogos-digitais>. Acessado em 29 de Novembro de 2016.
10. Steam Community: GestureWorks Gameplay. <https://steamcommunity.com/app/296610>. Accessed 29/11/2016.
11. Hassenzahl, M. 2005, Hedonic, emotional, and experiential perspectives on product quality. In Encyclopedia of Human Computer Interaction, pp. 266-272.
12. Nacke, L., E., Grimshaw, M., N., Lindley, C., A., 2010, More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. Interacting with Computers, Springer, pp. 336-343.
13. Mandryk, R. L., Inkpen, K. M., & Calvert, T. W. 2006. Using psychophysiological techniques to measure user experience with entertainment technologies. In Behaviour and Information Technology. vol. 25 no. 2, pp. 141-158. DOI=10.1080/01449290500331156.
14. Mirza-Babaei, P., Nacke, L. E., Gregory, J., Collins, N., Fitzpatrick, G. 2013. How Does It Play Better? Exploring User Testing and Biometric Storyboards in Games User Research. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1-10.
15. Nacke, L., E. and Lindley, C. 2009. Affective Ludology, Flow and Immersion in a First- Person Shooter: Measurement of Player Experience. The Journal of the Canadian Game Studies Association. Vol 3, No 5, pp. 1-21.
16. Ramirez, R. and Vamvakousis, Z. 2012, Detecting Emotion from EEG Signals Using the Emotive Epoc Device. In Brain Informatics, Springer. pp. 175-184.
17. Cernea, D., Kerren, A., and Ebert, A. 2011. Detecting Insight and Emotion in Visualization Applications with a Commercial EEG Headset. In Conference on Evaluations of Graphics and Visualization - Efficiency, Usefulness, Accessibility, Usability.
18. Badcock, N. A., Mousikou, P., Mahajan, Y., De Lissa, P., Thie, J. and McArthur, G. 2013, Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs. PeerJ.
19. Emotiv: Epoc Brain Activity Map. <https://www.emotiv.com/product/epoc-brain-activity-map/>. Access 18/11/2016.
20. Choppin, A. 2000. EEG-based human interface for disabled individuals: Emotion expression with neural networks. Master's thesis, Tokyo Institute of Technology, Yokohama, Japan.
21. Emotiv. <https://www.emotiv.com/> Access 18/11/2016.
22. Emotiv: User Manual For Emotiv Xavier Control Panel Epoc. <https://www.emotiv.com/>. Access 18/11/2016.
23. Schrepp, M., Held, T. and Laugwitz, B. 2006. The influence of hedonic quality on the attractiveness of user interfaces of business management software. In Interacting with Computers, 18, 5, pp. 1055-1069. DOI=10.1016/j.intcom.2006.01.002.
24. Sauro, J. and Dumas, J. S. 2009. Comparison of Three One-Question, Post-Task Usability Questionnaires. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1599-1608.
25. Lankes, M., Bernhaupt, R., Tscheligi, M. 2010. Evaluating User Experience Factors Using Experiments: Expressive Artificial Faces Embedded in Contexts. Evaluating User Experience in Games. Springer London. p.165-183. DOI=978-1-84882-963-3.
26. Brown, M. A., Kehoe, A., Kirakowski, J. and Pitt, I. 2015. Beyond the Gamepad: HCI and Game Controller Design and Evaluation. In Game User Experience Evaluation, Springer, pp. 263-285.
27. AttrakDiff. <http://attrakdiff.de/index-en.html>. Access 08/12/2016.
28. Brooke, J. 1996. SUS-A quick and dirty usability scale. Usability evaluation in industry, v. 189, n. 194, p. 4-7.
29. Bernhaupt, R. and Pirker, M. 2013. Evaluating User Experience for Interactive Television: Towards the Development of a Domain-Specific User Experience Questionnaire. In Human-Computer Interaction – INTERACT 2013, Springer, pp. 642-659.
30. Pelegrino, M. 2016. Um Controle Virtual Dinâmico para Jogos Adaptado ao Gameplay e ao usuário. Dissertação de Mestrado - IC - UFF 2016.
31. Arnold, A. 1999 Mental effort and evaluation of user interfaces: A questionnaire approach. In: Proceedings of 8th International Conference on Human-Computer Interact, ACM, New York, pp. 1003–1007.
32. McNamara, N. and Kirakowski, J. 2011. Measuring user-satisfaction with electronic consumer products: The Consumer Products Questionnaire. In: International Journal of Human-Computer Studies, 69, 6, pp. 375-386. doi=10.1016/j.ijhcs.2011.01.005.
33. Flanagan, J. C. 1954. The Critical Incident Technique. Psychological Bulletin. pp. 1-33.
34. Nacke, L. E. 2013. An Introduction to Physiological Player Metrics for Evaluating Games. In Game Analytics: Maximizing the Value of Player Data. pp 585-619. doi=10.1007/978-1-4471-4769-5\_26.
35. Bradley, M. M., & Lang, P. J. 2007. Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), Handbook of psychophysiology (3rd ed., pp. 581–607). New York: Cambridge University Press.