

An Evaluation of the Impact of Game Genre on User Experience in Cloud Gaming

Peter Quax, Anastasiia Beznosyk, Wouter Vanmontfort, Robin Marx and Wim Lamotte

Hasselt University - tUL - iMinds
Expertise Center for Digital Media
Wetenschapspark 2
3590 Diepenbeek, Belgium
Email: peter.quax@uhasselt.be

Abstract—Cloud gaming, in which the processing power of a datacenter-based infrastructure is utilized versus local resources, is a popular topic in research. This technology is successfully applied by vendors to enable low-end hardware to deliver a similar gameplay experience to state of the art consoles. Many works in literature have focused on the quantitative aspects of the technology (i.e. delay measurements, visual quality determination etc), but the qualitative factors have not received a similar systematic treatment. Games are typically classified in terms of their gameplay into a distinct category or genre, including action, puzzle, strategy and racing games. In this work, a qualitative comparison of these genres is presented based on a common testing methodology which combines both objective (based on physiological measurements) and subjective (based on user evaluation) approaches. While in normal networked games, only multiplayer experiences are subject to the detrimental effect of delay, the nature of cloud gaming may result in an impact on singleplayer experiences as well. Results from this analysis hint at the fact that there is a similarity in delay-sensitiveness over the different genres in both singleplayer cloud gaming setups and traditional networked multiplayer games. More in particular, results show that action-oriented games are more sensitive to network delay in both setups when compared to other genres.

I. INTRODUCTION AND RELATED WORK

Cloud gaming services such as OnLive, GaiKai and Core Online have recently triggered interest in the gaming community by allowing low-end devices (i.e. cheap set top boxes or mobile devices) to stream games that were typically reserved for more high-end consoles. Besides cost-efficiency and power consumption, platform independence is a major advantage of these solutions. This is demonstrated by the fact that Sony is likely to deliver compatibility with PS3 games on its upcoming PS4 console using GaiKai technology [1]. The client software required for connection to these online services is also being migrated to the browser, delivering even more flexibility with regards to target platforms [2].

The major drawback of this solution – besides the huge investment costs in terms of hardware in the data center and associated scalability issues – is the fact that all data has to be streamed to and from the client. Each interaction that is performed is therefore subject to the round trip delay (RTT) between the end user and the data center. Although the back-end systems are optimized to minimize the processing delay within the data center, the access link (which is not under control of the gaming service provider) is an important

contributor to the final RTT figure and its associated access link delay cannot be eliminated. Given the expensive nature of data center facilities, access to these services is often limited to specific geographical regions where games can be played. As an example, OnLive has a European infrastructure located in data centers in the UK and Luxembourg and only allows (paid) accounts to be registered from locations within the UK, Luxembourg and Belgium. Cloud gaming providers have also opted to show warning messages if delay values exceed a pre-determined amount. Lots of studies have focused on determining the final delay experienced by the user [3], [4], [5] and to some extent also the influence on the experience [6], [7], but only few reports have focused on the relationship to the game subject matter [8].

Lots of research has been carried out quantifying the influence of increasing delay on user experience and user performance (subjective and objective) in traditional networked multiplayer games [9]. We will refer to the latter as the classical scenario. Note that these are always multiplayer games, as this is the only type in which the network plays a role in the determination of what is visualized on the screen and when.

Typically, one can conclude from existing literature on the classical scenario that games that exhibit an important action element (i.e. they are considered interaction-heavy) are more subject to negative influences of an increased delay. Narrowing down the entire range of games into 4 representative genres, one can summarize that action(1) and racing(2) games are typically more sensitive than puzzle(3) or strategy(4) games. In FPS (action) games, player interactions require intricate synchronization over the network to determine e.g. whether a target was hit or if a certain item was picked up by a player. A similar thought process is behind the classification of racing games as delay-sensitive due to the occurrence of collisions between players and its end-effect on the (local) simulation of the car dynamics. RTS games on the other hand feature lots of interactions of which the outcome can be computed locally and which do not need to be synchronized to the same degree as in an FPS - e.g. micro-management of units can be accomplished locally and update rates are limited to the speed at which the player can edit settings for a specific unit. The same goes for puzzle games, in which typical interactions consist of clicking an item and waiting for a return action from the opponent (i.e. turn-based).

One of the main distinguishing factors between networked games played in the classical scenario and those played on a cloud gaming infrastructure is that, in the latter, all user actions are subject to the RTT delay. This includes all movements of the mouse or key presses on the keyboard. In fact, no user actions can be visualized locally, but rather have to be interpreted server-side first and then visualized in a video frame that is sent back to the client. Therefore, one might intuitively think that all game genres would be impacted equally by the presence of delay. Even more, puzzle games might be considered as a more challenging scenario for cloud gaming versus action games, as very intricate (small) mouse movements may need to be interpreted which would otherwise be visualized locally (without network and processing delay). At the same time, there is a clear distinction between the types of games that are influenced by delay in both scenarios. In the classical scenario, only multiplayer games were subject to such impact, while in the cloud gaming scenario this is also the case for singleplayer games. In fact, in the cloud gaming scenario, the impact on multiplayer games might be lessened as the synchronization between players is no longer impacted by the distance between the game server and client. In the cloud scenario, the client is virtually located within the data center and thus has an (in theory) optimal connection to the game server at all times. It is also the reason why only singleplayer games are discussed in this paper; the impact of the cloud gaming setup itself (i.e. the delay associated with the visualization of each and every action, rather than the multiplayer aspect) is under scrutiny.

It is the goal of this work to compare the 4 game genres that were also studied in the classical scenario. Remember that in those cases, the game genres could be clearly divided into two classes : more delay-tolerant ones like puzzle and strategy games and less tolerant ones such as action and racing games. More specifically, the objective in this work is to find out – through an experimental approach – to what degree the presence of delay has an influence on the user experience and enjoyment and to see whether these findings align with the classical scenario. It was already mentioned earlier in this section that, intuitively, it is not certain whether these findings would apply to both scenarios.

Combining different metrics is important to get a thorough understanding of user experience – which is why multiple methods were utilized. Besides the widely-used subjective evaluation through a questionnaire, two methods to obtain objective data (i.e. measuring galvanic skin response (GSR) and pressure exertion on input device) have been employed. An analysis of these measurements may indicate changes in the (emotional) player state triggered by the gaming conditions. These physiological measurements were obtained through a special setup that made sure that data was collected through non-obtrusive means by removing the need for wired sensor connections between the user and the measurement setup.

II. TEST SETUP AND PROCEDURE

The results presented in this paper are based on the OnLive cloud gaming platform. This commercial service is available in Belgium and has a data center in Luxembourg. For the purpose of this test, a subscription was arranged to make sure that the quality of the service delivered would not be impacted

Genre	Game Title
Action	Oddworld: Stranger's Wrath
Racing	Little Racers Street
Strategy	Fate of the Pharaoh
Puzzle	World of Goo

TABLE I: Game selection within genre classification.

by (possibly) lower network/processing priorities assigned to free trial accounts.

A. Game selection

As indicated in section I, four representative game genres were chosen for investigation. A specific game was chosen within each of the categories, taking into account the fact that a stand-alone version of the game was required for the training phase of the experiment (the reason behind this is discussed later). The games in table I were ultimately selected for the experiment.

B. RTT value selection

In a first stage, the standard (network) round trip time had to be determined. A Wireshark trace of an OnLive session results in a set of IP addresses that geolocate to the data center in Luxembourg; however the specific machines do not respond to ICMP messages and can therefore not be ping'ed. An estimation of the network RTT was made by taking an average of the ping times to several web sites hosted in Luxembourg from the actual experiment location. As the experiment was performed on the premises of the Expertise Center for Digital Media, which is directly connected to Belnet (the research network in Belgium) and in turn to other European research networks, typical delay values are below 5 milliseconds. Although it cannot be calculated or determined, the processing delay of the cloud gaming infrastructure itself can be considered near-constant when playing during off-peak hours (i.e. between 9AM and 5PM). There is existing literature which estimates these values. For the remainder of the discussion, we will assume these inherent delays are constant and are low enough (compared to the chosen test values) to be disregarded for reasons of clarity.

The baseline condition (=most optimal case) in which the system can be used is assumed to be that of a direct connection between the lab network and the Luxembourg data center (with a typical RTT of less than 5ms). To simulate varying conditions (e.g. a user located in a more remote location with respect to the data center or a user on a wireless network link), the delay values needed to be artificially increased and controlled. While the next section describes how this is performed in practice, a determination of appropriate values has to be made first. In literature, one can find evidence that a delay value of 100ms seriously impacts player behavior and experience in the least delay-tolerant genre, i.e. first person shooter games - in the traditional (non-cloud) scenario. Again, note that this solely concerns multiplayer games. At the same time, playing on the OnLive infrastructure with such delay values (100ms RTT) will result in a system warning that optimal gameplay cannot

be guaranteed. This scenario (100ms additional RTT delay) is therefore considered to be the worst case scenario. Prior empirical evidence (obtained before the actual experiment) also indicated that values beyond 100ms result in near-unplayable conditions and would skew the results too heavily. Also of interest is the question whether a gradual increase of the impact can be observed. Therefore, a midway value of 50ms was chosen as a third test condition. In the traditional scenario, only the most delay-sensitive games are subject to an impact under these conditions.

C. Physical setup & methodology

As indicated in section II-B, a setup was used that enables systematic testing of gameplay under varying network conditions. Technically, this was achieved by connecting a PC to the Belnet network infrastructure through a Linux NetEM bridge. This allows for continuous tweaking of parameters such as delay and packet loss. In this case, a virtual Ethernet bridge adapter was configured (based on two physical NICs), whereby half of the RTT value was assigned to each individual physical adapter in order to be able to introduce delay in both ways. To rule out any influence of other factors, the (additional) packet loss value was set to 0.

The experiment was conducted in two phases. First, the participants (8 people from the research lab, varying in age between 22 and 30) were able to get familiar with the games through a preliminary warm-up session. For this session, the games were played locally (which shows why it was important in the selection of games to make sure that an off-line version was available as well) on the same hardware used for OnLive testing – to rule out any impact of the physical appearance of the hardware/software and visual (generated) on-screen output on the user experience. Either one or two initial complete levels of each game were assigned to the players during this preliminary session, or a time limit was set for those games with lengthier levels (e.g. the action game). The purpose of this training session was to rule out possible learning effects during the actual OnLive experiments by making sure that all players were equally familiar with the games. At this stage, the level of skin conductivity was measured under normal gaming conditions (baseline). These figures were subsequently used to determine the change in emotional player state when subjected to delay. Additionally, to find out whether players exert additional force on input devices when subject to a certain degree of delay (e.g. pushing down on keys with increased force in case of frustration or stress), a maximum pressure exertion in a neutral state was captured by asking the player to hold down a button for about 30 seconds with typical force.

The actual experiment, which took place the day after the training sessions and involved the same participants, consisted of four gaming sessions – one for each genre – of about 12 minutes. During each session, the delay value was adjusted every 2 minutes in a random way (out of the possible values of 0, 50 and 100ms RTT, making sure that each value was applied twice during the session). Care was taken to make sure that each player was subject to the various conditions in an equal amount (but at different times throughout each game). All games were played consecutively by all players, who were not informed about the purpose of the test or the changes to the network that would happen during gameplay.

As mentioned in the introduction, to make a distinction between the actual influence of the network quality (objectively) on player performance (objectively) versus the user perception (subjectively), a combination of different physiological measurements (GSR and pressure exertion) were obtained during gameplay testing to see whether an impact was clear. The setup was carefully designed to be non-obtrusive to the user, which is in contrast to most experiments that include physiological measurements where the user is connected to the recording devices through a set of wires, restricting movement and making him/her constantly aware of being 'probed'. The setup used for this experiment consisted of both a pressure sensor and a GSR recording using a custom sensor. The pressure sensor was mounted on the left button of the mouse (for action, puzzle and strategy games) or on the 'up' key of the keyboard (for the racing game). Similarly, the GSR sensor was mounted to follow the curvature of the mouse for action, puzzle and strategy games and on the keyboard palmrest for the racing game. A time-synchronized log was maintained that could be referenced with the changes in delay value (which happened every 2 minutes). To quantify the subjective experience, users were asked to 'think aloud' while playing the game, and also to fill in a questionnaire after completion of all four sessions. The latter option was chosen after careful deliberation, taking into account the total duration of each test run (for the participant) and the fact that it is – in general – not desirable to interrupt players whilst performing the test. A similar approach was taken in earlier work undertaken by the authors. Seeing as the 'think aloud' approach did not yield many interesting results (probably due to the concentration required while playing the game), the questionnaires proved to be the main way to obtain information on the subjective experience.

III. RESULTS

A. Objective observations

To see whether an objective conclusion could be drawn based on the physiological measurements, which might indicate increased levels of intensity in game play or frustration, the sensor values were compared between the three delay conditions (baseline, 50ms and 100ms RTT). Firstly, the pressure exertion on input devices was analyzed. Similarly to [10], each pressure value was scaled to an index between 0 and 1, based on the maximum value obtained during the preliminary session. As has been shown in earlier studies [11], similarly to other physiological signals, an increased pressure index can be associated with user frustration. The goal of this setup was to determine whether an increase in delay would result in increased frustration. Although in general, these tests did not reveal truly surprising results, a general trend can be observed from the charts hinting at increased pressure being applied to the left mouse button in the action game with growing delay values - similarly but to a lesser degree in the puzzle game. The absolute values of the racing game cannot be compared directly to the others, due to the fact that the former was based on the pressure exerted on the 'up' key of the keyboard (which, in practice) is held nearly constantly down). These effects are visualized in figure 1.

Additionally, values obtained from the GSR sensor were analyzed for every player to determine whether or not any trend could be observed. The results (in absolute figures) are shown

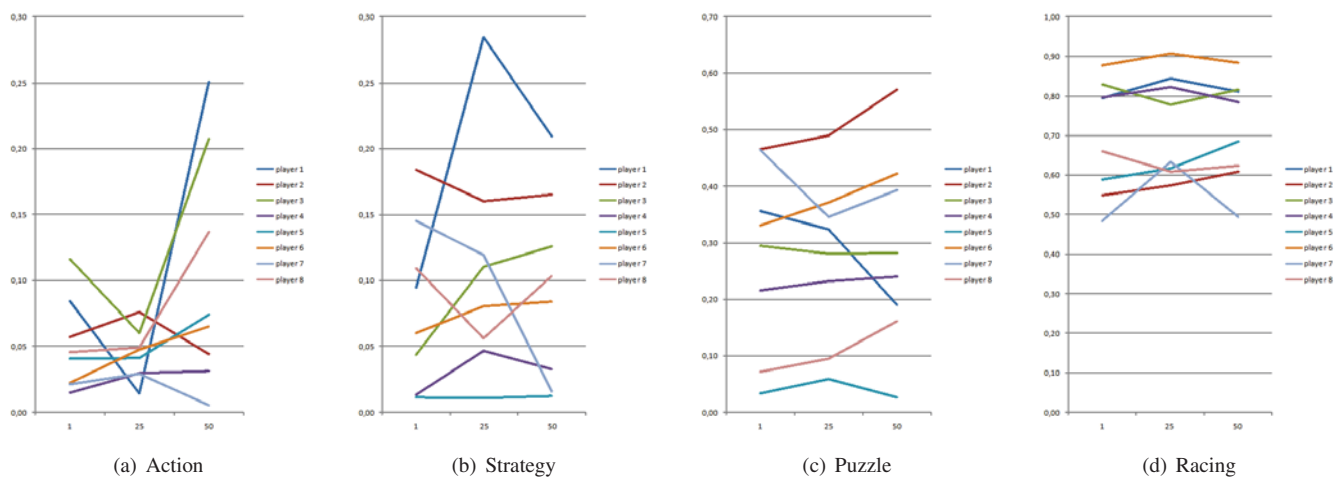


Fig. 1: Comparison of pressure applied in different games under three delay conditions.

in figure 2. In these charts, the values for test person 2 were omitted because of the lack of valid data points - which was due to difficulty experienced by this particular user in keeping his hand firmly on the top surface of the mouse). Additionally, for each test condition, an analysis was made to determine for how long (time, in percentage equivalent) every player experienced a higher skin conductivity than his/her baseline condition. The baseline is calculated as an average GSR level during gameplay without network degradation – i.e. during the preliminary test. Although more detailed analysis is clearly required to determine whether or not players were actually frustrated during this increase (rather than being influenced by other external factors), its presence indicates some changes in their emotional state (being other than normal).

The information obtained through the GSR sensor yielded similar results to the pressure setup. The graphs (see figure 3) demonstrate a more pronounced duration of time during which the player state differed from his/her baseline for the action game. Similarly to the analysis of the pressure indexes, in the racing game, players were almost constantly in an active playing state which is reflected in their GSR level – being very high. Although an analysis of both measurements have indicated some general trends when comparing game genres and differences between the no-latency and high-latency scenarios, neither of the two hinted at a (substantial) difference between the 50 and 100ms scenarios.

B. Subjective observations

More interesting results were obtained by analyzing the answers to the questionnaires that were filled out at the end of all four gaming sessions.

1) *Perceived network quality*: In a first analysis, it would be interesting to see whether the players noticed the changes in network quality differently while playing the 4 distinct game genres. Note that it is not the intent to see here whether a specific network condition (e.g. 50ms of RTT delay vs 100ms) would be noticed more explicitly, but rather a general trend between the genres. A specific question in the questionnaire asked the players to quantify this effect : "To what extent

did you perceive a negative influence of the changing network quality?"

First, an RM ANOVA test reveals that there is a significant difference between the four games ($F(3, 21) = 9,70$; $p < 0.01$). Further testing using a post-hoc LSD test revealed that in the puzzle and strategy games, the answer values to this question were significantly lower than in the action and racing games. These results align with those from classic networked games, in which the same two genres are also less impacted by delay. However, the reader is reminded that the games under investigation here are singleplayer experiences rather than multiplayer in the classical setup.

2) *Impact of network quality on enjoyment*: Secondly, a question was included to probe at the degree to which (a degraded) network quality impacts the user enjoyment between the different genres. In the questionnaire, the item "To what extent did the network quality negatively influence the level of your enjoyment?" queried the user about his/her findings. Again, a RM ANOVA test indicates that there is a significant difference between the four games ($F(1.35, 9.46) = 6.044$; $p = 0.004$). Using the same LSD post-hoc test as in section III-B1, the observation can be made that in the puzzle and strategy games, the degrading network quality has a significantly lower influence on player enjoyment versus the racing game. Analogously, in the strategy game, the network quality has a significantly lower influence on the enjoyment than in the action game. Again, it is apparent that the game genres that are typically more sensitive to delay also come out on top here.

3) *Impact of network quality on frustration*: To cross-check the results from the previous question, a third item in the questionnaire hinted at the experienced frustration by the user by asking : "To what extent did the network quality influence the level of your frustration?". Again, just like the previous results, a RM ANOVA test showed that a significant difference was present between the four games ($F(1.49, 10.39) = 9.21$; $p < 0.01$). This time, the LSD post-hoc test concluded that in the puzzle, strategy and action game, a degrading network quality had a significantly lower impact on player frustration versus the racing game. Similarly, in the

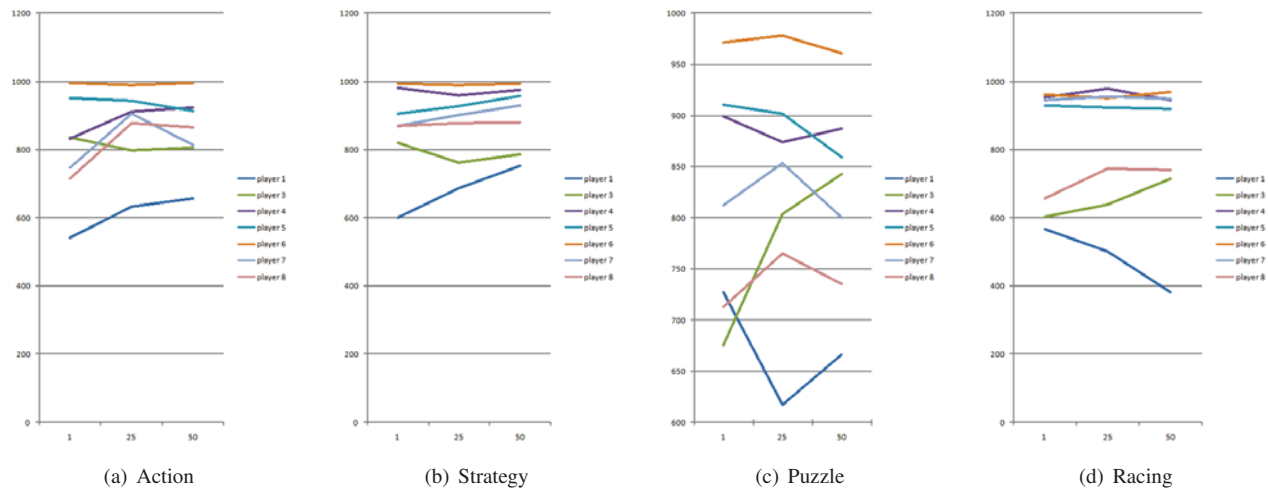


Fig. 2: Comparison of absolute GSR values in different games under three delay conditions.

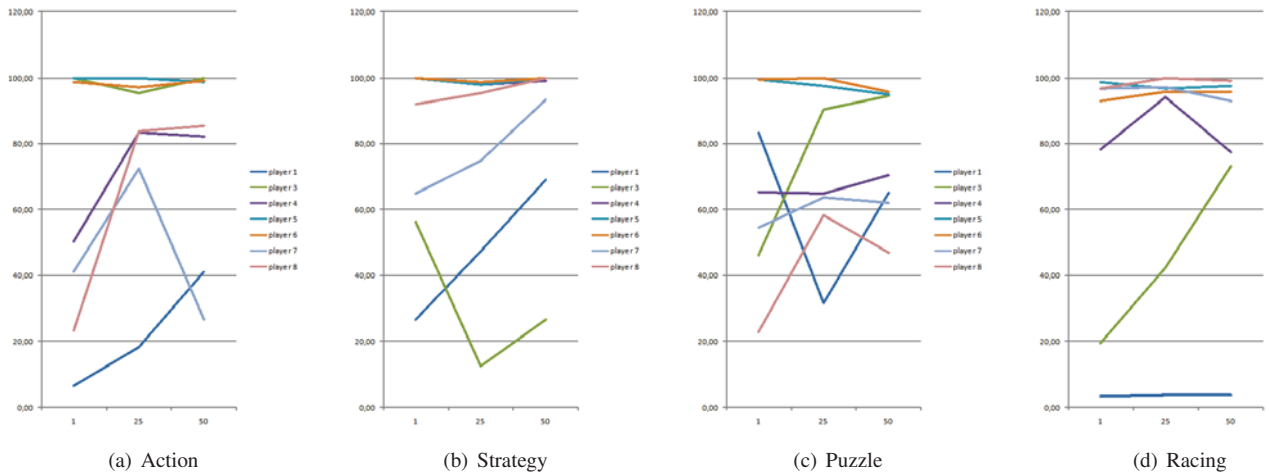


Fig. 3: Comparison of duration (in percentage) of the GSR value being over an individual user's baseline condition in different games under three delay conditions.

strategy game there was a lower impact compared to the action game. This confirms the results obtained from the previous question.

C. General observations

Informal oral discussions with the test subjects after the experiment showed that the impact of the increased delay (up to 100ms of additional RTT delay) was never noticed as particularly disturbing. The results from the analysis of the questionnaires however shows that these (apparently minor) side-effects are perceived differently for the various genres. This is also represented in figure 4. It should be remarked that the observations made above can only be used as an indication; further testing would be required for a more comprehensive insight into possible issues/contributing factors and threshold determination. It is also obvious that unobtrusive physiological measurements do not lead to the same (or even any) concrete conclusions. Additional research would be required to get

a better understanding of an appropriate way to objectively quantify the impact other than through questionnaires.

It is by no means certain that the games that have been chosen are representative for an entire genre within the chosen game classification taxonomy. However, the fact that a consistent result was obtained by analyzing the questions from the questionnaire, shows that there might be need for a different treatment of the games according to their genre in the back-end. This could in turn also be used on a more technical level to e.g. design QoS mechanisms that take into account this classification when prioritizing certain traffic flows in order to minimize end-to-end-delay. Cloud gaming providers might also be able to use such a classification to show appropriate warnings at different delay levels (the current system used by OnLive seems to be a binary system that applies the same warning threshold to all games, regardless of genre).

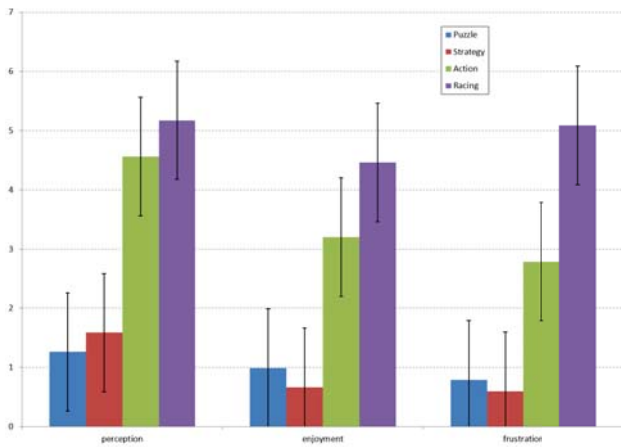


Fig. 4: Mean and standard deviation (error bars) values plotted over the different genres, for general perception of influence, influence on enjoyment and frustration respectively. Lower values indicate 'better' results.

IV. CONCLUSION AND FUTURE WORK

In this paper we have described an experiment to evaluate whether games played on a cloud gaming infrastructure are impacted to a different degree when faced with network delay based on their genre. It was found that, analogous to the classical (multiplayer) scenario, those games that require intense interactions like action and racing games were more sensitive to the presence of delay versus puzzle and strategy games. An important remark is that the results presented here are based on a distinct selection of example games which may or may not be generalizable to an entire genre within the chosen game taxonomy and that the evaluation is based primarily on questionnaires. However, the fact that at least some meaningful conclusions could be drawn might encourage others to perform further, more detailed, tests and to determine e.g. appropriate delay threshold values for cloud gaming infrastructures for various game genres. Additionally, it would be interesting to tackle the problem from a game design point of view and identify exactly which gameplay elements are most sensitive to delay; in a later stage these could be altered/adapted to optimize the user experience.

ACKNOWLEDGMENT

This work is part of the PIM Omega Project, funded by IWT Flanders. Thanks go out to the other partners in the project : Visual Impact, Androme, Neopica and GriN.

REFERENCES

- [1] IGN. (2013, Feb) Gaikai integrated with playstation 4. [Online]. Available: <http://www.ign.com/articles/2013/02/20/playstation-cloud-revealed>
- [2] OnLive. (2012, Jun) Onlive offers one-click in-browser cloud gaming. [Online]. Available: http://games.onlive.com/corporate/press_releases/1353/print
- [3] K.-T. Chen, Y.-C. Chang, P.-H. Tseng, C.-Y. Huang, and C.-L. Lei, "Measuring the latency of cloud gaming systems," in *Proceedings of the 19th ACM international conference on Multimedia*, ser. MM '11. New York, NY, USA: ACM, 2011, pp. 1269–1272. [Online]. Available: <http://doi.acm.org/10.1145/2072298.2071991>

- [4] S. Choy, B. Wong, G. Simon, and C. Rosenberg, "The brewing storm in cloud gaming: A measurement study on cloud to end-user latency," in *Network and Systems Support for Games (NetGames), 2012 11th Annual Workshop on*, 2012, pp. 1–6.
- [5] S. K. Barker and P. Shenoy, "Empirical evaluation of latency-sensitive application performance in the cloud," in *Proceedings of the first annual ACM SIGMM conference on Multimedia systems*, ser. MMSys '10. New York, NY, USA: ACM, 2010, pp. 35–46. [Online]. Available: <http://doi.acm.org/10.1145/1730836.1730842>
- [6] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hossfeld, "An evaluation of qoe in cloud gaming based on subjective tests," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on*, 2011, pp. 330–335.
- [7] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hossfeld, "Gaming in the clouds: QoE and the users' perspective," *Mathematical and Computer Modelling*, Dec. 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.mcm.2011.12.014>
- [8] Y.-T. Lee, K.-T. Chen, H.-I. Su, and C.-L. Lei, "Are all games equally cloud-gaming-friendly? an electromyographic approach," in *Proceedings of IEEE/ACM NetGames 2012*, Oct 2012.
- [9] A. Steed and M. F. Oliveira, *Networked Graphics: Building Networked Games and Virtual Environments*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2009.
- [10] J. Octavia, "Adaptivity in virtual environments: Enhancing user interaction and accommodating user diversity," Ph.D. dissertation, Hasselt University, Belgium, 2011.
- [11] W. van den Hoogen, W. IJsselsteijn, and Y. de Kort, "Exploring behavioral expressions of player experience in digital games," in *Proceedings of the workshop for facial and bodily expressions for control and adaptation of games (ECAG 2008), September 16 2008, Amsterdam, A. Nijholt and R. Poppe, Eds., ECAG 2008*. Amsterdam: ECAG 2008, 2008 2008, pp. 11 – 19. [Online]. Available: <http://repository.tue.nl/667606>