4th International Workshop on Perceptual Quality of Systems (PQS 2013)
2-4 September 2013, Vienna, Austria

Factors Influencing Gaming QoE: Lessons Learned from the Evaluation of Cloud Gaming Services

Sebastian Möller [1], Dennis Pommer [1], Justus Beyer [1], Jannis Rake-Revelant [2]

[1] Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin, Germany
[2] Telekom Innovation Laboratories, Deutsche Telekom AG, Germany

sebastian.moeller@telekom.de, dennis.ben.pommer@gmail.com, justus.beyer@telekom.de, jannis.rake-revelant@telekom.de

Abstract

In this paper, we present initial results from the evaluation of two computer games provided by a prototypical cloud gaming service provider. An experiment is described which addresses the impact of transmission bandwidth and delay on different aspects of user-perceived quality. It is analyzed with respect to the impact of the experimental procedure and set-up. The analysis shows that the set-up of the local player software and the restriction of the transmission channel considerably impact the subjective judgments. In addition, game content impacts performance metrics as well as perceived quality. Implications of these findings for the experimental set-up of cloud gaming quality evaluation are discussed.

Index Terms: gaming, Quality of Experience (QoE), evaluation methods, cloud services.

1. Motivation and introduction

Computer games become increasingly popular and – in case they are implemented on remote platforms and/or played at distributed locations – are on their way to significantly contribute to internet traffic. A possible way to implement computer games is in terms of a cloud service, i.e. the game itself resides in the cloud, while the user only has a physical device for controlling input and rendering output at her premises. In such a scenario, the quality experienced by the user during the game does not only depend on the game service itself (including the game logic, game design, interaction possibilities, etc.), but also on the performance of the transmission channel involved.

User-perceived Quality of Experience (QoE) when interacting with computer games differs in several respects from QoE of transmitted media or from interacting with task-oriented services. In contrast to pure media consumption, the user of a computer game interacts with the service, thus the timely and error-free delivery of information is critical for the gaming process. In contrast to task-oriented services (such as information or reservation web sites), the user of a computer game does her activity because the activity is rewarding in itself. Thus, there is no need for a quick termination of the process, and concepts as effectiveness and efficiency lose their positive value for overall quality. In turn, there are other quality aspects which need to be considered, such as positive and negative emotions resulting from playing, or the flow experienced by the user. A taxonomy of quality aspects and related metrics is provided in [1].

In order to quantify such quality aspects in a valid and reliable way, interaction experiments with human test participants are necessary. During such an experiment, participants interact with the gaming service to-be-evaluated in a more-or-less controlled way, and rate perceived quality after each interaction on dedicated questionnaires. In order to have full control over the experimental factors, the experiment is best being set up in a laboratory environment, as described e.g. in [2]. However, such a set-up is not always possible when real cloud games are to be evaluated; in such a case, it may be necessary to access the remote game platform via the internet, and to control just the environment of the playing participant. Whereas standardized methods for carrying out subjective quality evaluation methods abound for speech [3], video [4], and task-oriented interactive services [5], no such standards are available for computer gaming in general, and for cloud gaming in particular.

In this paper, we report on experiences made during the evaluation of a cloud gaming service. The service was offered by a commercial provider and consisted of several games which were offered, and a player software to be installed at the user side which captures the user input in terms of game commands, and visualizes and auralizes the game output (audio and video) on the user's computer. Thus, only commands and rendered audio and video were sent across the connection. We evaluated the service in a laboratory experiment with 19 participants, where we deliberately limited the maximum transmission bandwidth of the connection between cloud game and user computer, and where we introduced delay in some of the test conditions. Whereas the quality judgments for the games are not of scientific interest (because they are limited to the particular games we evaluated), our analysis focusses on the experimental factors which affected the obtained quality judgments.

The paper is organized as follows. In Section 2, we briefly review methods for evaluating gaming QoE. The experimental set-up is described in in Section 3, followed by the test procedure in Section 4. Section 5 analyzes the results with respect to the impact of the test set-up and procedure. The results are discussed with respect to the implications for future gaming QoE evaluations in Section 6, and conclusions are drawn in Section 7.

2. Gaming QoE evaluation methods

When evaluating the QoE of gaming activities, it is first important to differentiate between influencing factors, interaction performance metrics, and user-perceived QoE aspects.

Influencing factors relate to the user (experience, playing style [6], motivation, age, gender, emotional status, etc.), the system (including server, transmission channel, player, game genre [7], game structure and rules), as well as the playing context (physical and social environment, extrinsic motivations for playing, etc.). Whereas influencing factors of the user are commonly analyzed with the help of debriefing questionnaires,

10.21437/PQS.2013-31

system- and context-specific factors can be identified from detailed descriptions of the game and the playing situation. Interaction performance relates to metrics which characterize the playing behavior on both the system and the user side; for the latter, it can be measured e.g. in terms of perceptual effort, cognitive workload and physical response effort, e.g. via questionnaires, whereas for the former metrics depend on the distribution of the activity between the interface software and device, the backend platform, and the game itself. In cloud gaming, the common scenario is that the player interacts through a somewhat "lightweight" interface software installed on her device, whereas the game logic is implemented and the video and audio output is rendered on the back end in the cloud; thus, the transmission channel performance between player and backend is crucial for user-perceived quality. Transmission channel performance measurement requires e.g. physical measurement of delay times, or packet-based measurement of channel throughput.

User-perceived QoE can be considered along several dimensions. One aspect which has frequently been investigated in the past was "playability" [8][9], although no agreement has been reached about what this really includes. In its narrower definition [9], playability relates to the degree to which all functional and structural elements of a game (hardware and software) enable a positive player experience to the gamer, and is that a prerequisite for positive experiences; we prefer to call this aspect "interaction quality". This is amended by the learnability and intuitivity of the game, which we subsume under the term "playing quality" in our taxonomy [1], and by the appeal (aesthetics and novelty) of the game.

Player experience can be considered as the degree of delight or annoyance of the gamer. To our knowledge, Poels et al. [10] defined the most comprehensive taxonomy of player experience, which consists of the sub-aspects challenge, control, flow, tension, immersion, competence, positive affect and negative affect. These seven sub-aspects can be measured with the purposely-built Game Experience Questionnaire (GEQ), see [11]. These metrics are interesting when the overall experience of the gamer is of interest, but they do not provide a diagnostic picture of the impact of transmission characteristics (e.g. delay and throughput) on aspects of the interaction quality (e.g. output quality, interactive behavior) which are highly relevant in a cloud-gaming scenario.

Thus, in the study reported below, we designed a new questionnaire for the given purpose. This questionnaire includes three aspects related to the interaction quality (video quality, audio quality, input sensitivity), one aspect related to playing quality (complexity of the game situation), and three player-experience-related aspects (overall quality, pleasantness, perceived value of the cloud-gaming service). We are aware that the use of a non-validated questionnaire has inherent problems because the validity and reliability of the used scales is not known, but hope that our empirical data can contribute to the future development of appropriate measurement instruments for diagnostically evaluating cloud-gaming services.

3. Experimental set-up

The aim of our study was to analyze the impact of transmission performance aspects on user-perceived QoE of a commercial cloud gaming service. This service runs on a remote backend server (i.e. in the cloud) and is only accessible to us via an internet connection from a local computer. Whereas the characteristics of the Internet connection to our laboratory cannot be guaranteed for, we tried to deliberately introduce limitations of the transmission channel in our laboratory premises.

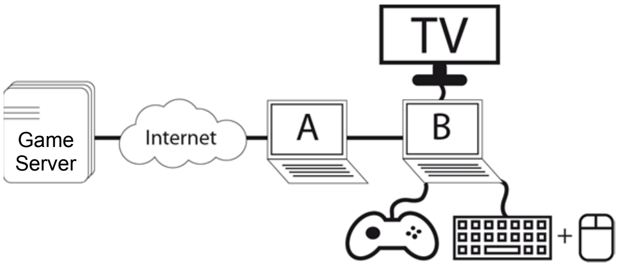

Figure 1: Experimental set-up.

The experimental set-up used for this purpose consists of two computers installed in a laboratory environment decorated like a living room, see Figure 1. The first computer (B) runs the player software and is connected to input (keyboard, mouse, Xbox controller) and output devices (HD video screen with 1920x1080 resolution, stereo loudspeakers). In addition, it is connected to the second computer (A) which provides the link to the remote cloud server running the game software via broadband Internet connection. The second computer also shapes the connection characteristics to the game server, in that it is able to limit the maximum data rate to 4.5 or 10 Mbit/s, and/or to introduce a delay of 50 or 125 ms to the packets, using a Linux-based traffic control queuing scheme. The maximum transmission rate was also limited by the available network connections to the Internet to 16 Mbit/s.

With the help of this set-up, two games (casual game and 3rd-person shooter) were tested under six different conditions (0, 50 or 125 ms delay combined with 16 and 4.5 (for the casual game) or 16 and 10 Mbit/s (for the 3rd-person shooter) transmission bandwidth. The games were selected as to provide different types of experiences to the gamer via their game genre and rules. In the casual game, the task of the gamer is to sheppard sheep to a spacecraft via several obstacles. This game is played with the help of the Xbox controller. The 3rd-person shooter involves moving avatars in the game, interacting with avatars, and fights. This game is played via the keyboard and mouse.

For each game, 7 different scenes have been selected as playing scenarios. The selection of scenes was guided by the idea to have comparable conditions for each transmission channel setting, and for each participant. Each scene lasted between 2 and 3 minutes and consisted of a brief description of a task or sequence of tasks which should be performed by the participant, like "Walk straight ahead until you reach the little square. On the square, you will be attacked by some enemies. Defeat them and walk to the fire place on the other side." Scenes have been selected to be comparable in length and difficulty, and to start from the same default conditions for each transmission channel setting. The first scene served as a tutorial and was not considered in the subsequent analysis.

4. Test procedure

19 participants which differed with respect to their experience in computer gaming (12 experienced gamers with >9h gaming activity per week, 7 casual gamers with less) took part in the test. Most of our participants were male (17), which resulted from the

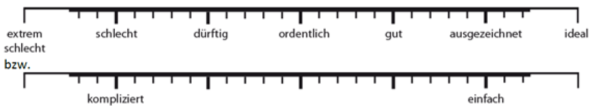

Figure 2: Rating scale layout for the overall quality scale (upper panel; labels: extremely bad, bad, poor, fair, good, excellent, ideal) and semantic differential scales (central panel; labels: complicated, easy). The lower panel illustrates the coding of ratings into values [0; 6].

recruiting from our (Technical University) student body. They were guided to a living-room lab environment, instructed about the purpose of the experiment, and filled in an introductory questionnaire. Then, they were instructed about the first game (random order) and fulfilled a tutorial, followed by six scenes from this game they had to play, each with a different transmission channel setting. After each scene, participants had to fill out a questionnaire judging the 7 mentioned quality aspects (input sensitivity, video quality, audio quality, overall quality, complexity of the game situation, pleasantness, perceived value) on continuous rating scales with 2 or 7 antonym labels, see Figure 2. This part was followed by the second game (1 tutorial and 6 scenes). The test was terminated with a debriefing questionnaire.

5. Result analysis

The analysis was carried out separately for each game, as the game had a significant effect on the results. The adjusted maximum bandwidth and delay were analyzed by a MANOVA with respect to their impact on the quality judgments, showing no significant interaction between these two factors.

Table 1: Mean values and standard deviations for the judgments of the casual game as a function of the adjusted transmission bandwidth.

Adjusted bandwidth	4.5 Mbit/s		16 Mbit/s	
Quality aspect	mean	std	mean	std
Input sensitivity	2.07	1.15	3.80	0.81
Video quality	2.99	1.08	3.52	1.05
Audio quality	3.77	0.81	4.17	0.76
Overall quality	2.88	1.11	3.91	0.72
Complexity	3.27	1.30	4.11	0.76
Pleasantness	2.93	1.23	4.14	0.98
Perceived value	3.05	1.11	3.95	0.75

For the casual game, the adjusted maximum bandwidth showed a significant negative impact on all 7 quality judgments (t-Tests, $p<0.01$ in all cases). For all judgments, the average ratings were more positive in the 16 Mbit/s than in the 4.5 Mbit/s case, see Table 1. However, we found that the adjusted maximum bandwidth did not always correspond to the real bandwidth in this game. In fact, only in rare cases transmission bandwidths above 5 Mbit/s have been observed for the casual game, even when the maximum bandwidth was set to 16 Mbit/s. For the high maximum bandwidth, we found a relatively high standard deviation of the bandwidth during the game, which was not observed when the maximum bandwidth was limited to 4.5 Mbit/s, see Figure 3. This led to a large standard deviation of the actual bandwidth for these conditions, which was neither observed for the 4.5 Mbit/s setting, nor for the 3rd-person shooter game (any bandwidth).

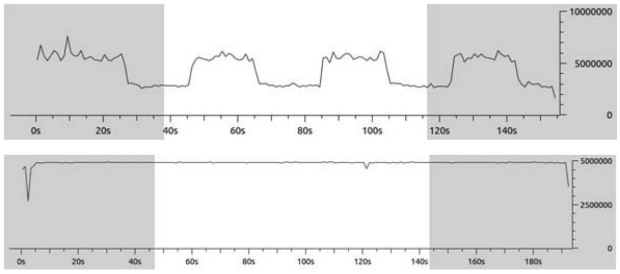

Figure 3: Actual bandwidth for two exemplary conditions of the casual game with 16 (upper panel) and 4.5 Mbit/s (lower panel) maximum bandwidth setting.

An analysis of the cases with high standard deviation showed that these cases were judged lower with respect to their video quality, and higher with respect to the input sensitivity. We expect that the player software installed on the user computer provokes a “balancing” of the traffic in case the maximum bandwidth is 4.4 Mbit/s. This behavior was not expected and not observed when we tested the set-up prior to the experiment, with an earlier player version.

For the 3rd-person shooter, we adopted a slightly higher restricted bandwidth (10 Mbit/s) than for the casual game, as we expected that the effect of bandwidth restriction would be more noticeable for this game, which was confirmed in our pre-tests. However, analyzing the effect of bandwidth restriction on the 7 quality judgments did not show a statistically significant effect.

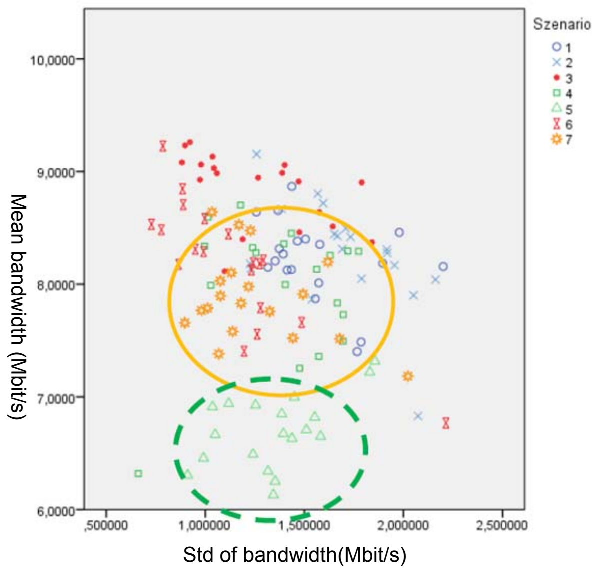

Figure 4: Mean bandwidth and standard deviation for different scenarios of the 3rd-person shooter game.

was a (statistically not significant in the t-test) trend towards a higher rating for the higher bandwidth (video quality: 3.87 (10 Mbit/s) vs. 4.19 (16 Mbit/s), p=0.058; audio quality: 4.01 (10 Mbit/s) vs. 4.32 (16 Mbit/s), p=0.056). A look at the real data rates showed that the maximum bandwidth did in no case exceed 9.4 Mbit/s, a value which was significantly higher in the pre-tests. We expect that the update of the player software between pre-test and test led to the reduced bandwidth requirement, and thus to no observable effect on the subjective judgments. Once again, this shows that the characteristics of the player software largely determine the subjective effects of transmission performance problems.

For the 3rd-person shooter, the actual bandwidth depended strongly on the scenario, see Figure 4. We expect that this is due to the degree of activity which differed between the scenarios. For example, scenario 5 which contained no fight scenes required a substantially lower bandwidth than the other scenarios, an effect which was not observed for the casual game. Thus, in addition to the type of game the scene which is used as a scenario for the test participants needs to be considered as an influencing factor when games with differing degree of gamer activity are to be evaluated.

The delay conditions did not show any significant effect for the casual game. For the 3rd-person shooter, there were statistical trends in two of the judgments (video quality and complexity of the game situation), but the latter in the opposite way compared to what we expected, namely that the delay condition was rated more favorably. We think that the effect for the complexity may be due to a less demanding game situation when delay was present. When considering only the judgments of the experienced gamers (>9h playing per week), there was a trend that the input sensitivity was judged less favorably with delay, but once again this was not statistically significant. No effect on gamer experience was observed for the casual game.

6. Discussion

The results show that the game set-up is particularly critical when cloud games are to be evaluated. A rather straight-forward limitation of the maximum transmission bandwidth might not fulfill the purpose, especially when the player software residing at the user's end counter-balances some of the effects of the restricted transmission channel. This problem can be circumvented by installing player software which is fully under control by the experimenter, but this is frequently not possible when commercial cloud gaming services are to be evaluated.

Another significant effect on the subjective judgments is given by the game itself. The scene which is selected for each experimental condition may either directly impact the subjective judgments, but also impact the necessary transmission bandwidth, especially when differing degrees of gamer activity are required within a single game. Thus, it is important that a variety of games and game scenes are used when evaluation results are expected to be representative.

Overall, we expect that gamer experience might have exercised an effect on some of the results. Still, to our knowledge there is no universally accepted assessment method for gamer experience which could be used for selecting corresponding test participants. Our classification scheme (experienced gamers with >9h of playing time per week, casual gamers with less) was motivated by the distribution of playing time of our participants, but might not be adequate for the purpose. We also tried to partly mitigate the effect of gamer experience by providing a training scenario, but this scenario might have been too short for achieving the desired effect. Whereas none of the participants knew the casual game, about half of them (53%) knew the 3rd-person shooter.

The distribution of participants in our study (17 male, 2 female) prevented us from performing gender analyses, thus we could not confirm differences which have been observed in other studies. Most of our participants were relatively young (mean 27.8 years) had a relatively good DSL internet connection at home (63% with ≥ 16 Mbit/s), which might have influenced their expectations, in particular with respect to reaction times. However, because of the limited participant pool, we could not perform any further analyses on these aspects.

The study described here addressed two games which are representative for a cloud gaming set-up. In this set-up, user actions are particularly vulnerable to transmission channel effects, as the commands from the user need to be transferred to the game backend. In addition, the video is rendered on the backend in the cloud and needs to be transferred to the user, making this part vulnerable to packet loss and coding artefacts (which were not addressed in our study). We assume that the effects might be different for a classical online game scenario. The results might also be influenced by the type of user interface (Xbox controller for the casual game, keyboard/mouse for the 3rd-person shooter), and might be different with other types of interfaces.

7. Conclusions

We presented initial results from a cloud gaming experiment in which the effect of transmission bandwidth and delay on seven aspects of gaming QoE were analyzed, taking two different games (a casual game and a 3rd-person shooter) as examples. The results were analyzed with respect to the impact of the experimental set-up and the game on user-perceived QoE. We found that the experimental set-up is critical in a cloud gaming context, as the implementation of the player software on the user's device may have a significant impact of the required transmission bandwidth, and hereby affect perceived QoE. In addition, the type of game and the particular scene selected for each trial may affect transmission bandwidth. Some effects seem to be limited to particular user groups (here experienced gamers), but the size of the experiment prevented us from further analyses.

In future experiments, we would like to analyze some of the observed effects in more detail, using a larger participant tool with a more diverse – and perhaps more representative – distribution of gender and expertise. In order to obtain more controllable conditions, we would like to develop a test set-up which leaves the player software as passive as possible, so that the effect of transmission channel degradations can be tested in a more reliable way. Ideally, the gaming QoE community would be able to agree on (a set of) standard set-ups which could be made available as open-source software, and which would allow for repeatable experiments. This way, data from different sources may be easily linked and pooled in order to obtain a better picture of the factors which affect cloud gaming QoE.

A similarly positive effect might come from standardized and validated questionnaires and potentially other (e.g. physiological) methods to capture different aspects of user-perceived QoE. The GEQ questionnaire might be a good starting point for this, but we deliberately chose to not use it in our study, because we thought that it would not cover the effects which are relevant for cloud gaming in a detailed way. This hypothesis could be checked further, and the questionnaire might be amended by others which would help to establish a toolbox of well-validated and useful metrics for cloud-gaming QoE, and for gaming QoE in general.

8. References

[1] Möller, S., Schmidt, S., Beyer, J., "An Overview of Concepts and Evaluation Methods for Computer Gaming QoE", submitted to Fifth Int. Workshop on Quality of Multimedia Experience 2013 (QoMEX 2013), AT-Klagenfurt, 3-5 July, 2013.

[2] Jarschel M., Schlosser D., Scheuring S., Hoßfeld T., "An Evaluation of QoE in Cloud Gaming Based on Subjective Tests", Workshop on Future Internet and Next Generation Networks (FINGNet 2011), Seoul, Korea, June 2011.

[3] ITU-T Rec. P.800, Methods for Subjective Determination of Transmission Quality, Int. Telecomm. Union, Geneva, 1996.

[4] ITU-T Rec. P.910, Subjective Video Quality Assessment Methods for Multimedia Applications, Int. Telecomm. Union, Geneva, 2008.

[5] ITU-T Rec. P.851, Subjective Quality Evaluation of Telephone Services Based on Spoken Dialogue Systems, Int. Telecomm. Union, Geneva, 2003.

[6] R. Bartle, Hearts, clubs, diamonds, spades: Players who suit muds, 1996, http://www.mud.co.uk/richard/hcds.htm.

[7] M.J.P. Wolf, "Genre and the video game". In M.J.P. Wolf, ed., The Medium of the Video Game, chapter 6, pp. 113-134, University of Texas Press, Austin TX, 2001.

[8] G. Sánchez, F.L. Gutiérrez, M.J. Cabrera, and N. Padilla Zea, "Design of Adaptive Videogame Interfaces: A Practical Case of Use in Special Education", in Proc. 7th Int. Conf. on Computer-Aided Design of User Interfaces (CADUI 2008), pp. 57-63, 2008.

[9] S. Engl, Mobile Gaming – Eine empirische Studie zum Spielverhalten und Nutzungserlebnis in mobilen Kontexten, Magister thesis, Universität Regensburg, 2010.

[10] 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 Proc. Futureplay 2007, Toronto, Canada, pp. 83-89, 2007.

[11] W. Ijsselsteijn, Y. De Kort, K. Poels, F. Bellotti, and A. Jurgelionis, "Characterising and Measuring User Experiences in Digital Games", in Proc. Int. Conf. on Advances in Computer Entertainment Technology, 2007.