Cloud Gaming QoE Models for Deriving Video Encoding Adaptation Strategies

Ivan Slivar
Faculty of Electrical Engineering and Computing
University of Zagreb
Zagreb, Croatia
ivan.slivar@fer.hr

Lea Skorin-Kapov
Faculty of Electrical Engineering and Computing
University of Zagreb
Zagreb, Croatia
lea.skorin-kapov@fer.hr

Mirko Suznjevic
Faculty of Electrical Engineering and Computing
University of Zagreb
Zagreb, Croatia
mirko.suznjevic@fer.hr

ABSTRACT

Cloud gaming has been recognized as a promising shift in the online game industry, with the aim being to deliver high-quality graphics games to any type of end user device. The concepts of cloud computing are leveraged to render the game scene as a video stream which is then delivered to players in real-time. Given high bandwidth and strict latency requirements, a key challenge faced by cloud game providers lies in configuring the video encoding parameters so as to maximize player Quality of Experience (QoE) while meeting bandwidth availability constraints. In this paper we address this challenge by conducting a subjective laboratory study involving 52 players and two different games aimed at identifying QoE-driven video encoding adaptation strategies. Empirical results are used to derive analytical QoE estimation models as functions of bitrate and framerate, while also taking into account game type and player skill. Results have shown that under certain identified bandwidth conditions, reductions of framerate lead to QoE improvements due to improved graphics quality. Given that results indicate that different QoE-driven video adaptation policies should likely be applied for different types of games, we further report on objective video metrics that may be used to classify games for the purpose of choosing an appropriate and QoE-driven video codec configuration strategy.

Keywords

Cloud gaming, QoE, Adaptation strategies, QoE modelling

1. INTRODUCTION

The cloud gaming paradigm is commonly characterized by game content delivered from a server to a client as a video stream, with game controls sent from the client to the server. The execution of the game logic, rendering of the 3D virtual scene, and video encoding are performed at the server, while the client is responsible for video decoding and capturing of client input. While such a game streaming paradigm significantly reduces the end client device requirements as compared to "traditional" online gaming and thus allows for the delivery of graphically-rich games to less powerful client devices, the downlink bandwidth requirements are significantly increased. Furthermore, gaming in general is a highly interactive service, thus imposing strict latency requirements. In the case of cloud gaming, meeting these requirements becomes very challenging (e.g., under 150ms of RTT is needed for good quality of First Person Shooter games [9]), with the need to calculate game state, render the virtual scene, and encode/decode the video stream. The available time budget (i.e., 150 ms) is used by network delay, virtual world state calculation, 3D scene rendering, video encoding on the server and decoding on the client side. There is thus not enough time for the server-side video encoder to optimize the bandwidth size of the sent video stream.

With available network resources varying over time, subject to issues such as varying access network conditions or a varying number of players accessing a bottleneck link, there is a need for efficient and dynamic service adaptation strategies on the game server to meet different bandwidth availabilities. A challenge faced by cloud gaming providers is configuration of the video encoding parameters used for game streaming with respect to different available network bandwidth conditions. As stated by Hong *et al.* [11], the cloud gaming server has no control over network latency, with packet loss and end-to-end delays resulting in lower effective bandwidth measured by the server. Hence, codec reconfiguration decisions made by the cloud gaming server (in terms of chosen target bitrate and framerate values) are driven by measured available effective bandwidth. Previous studies have shown that the codec configuration strategy is dependant on different game types [11, 20], with one possibility of classifying game genres being according to similar video and gameplay characteristics [14].

In this paper we build on both the results reported in [11] and in [20] with the aim of specifying video encoding adaptation strategies applicable in the context of cloud gaming. Hence, the problem we are looking to address is how to adapt the video encoding parameters of the game video stream in light of decreased bandwidth availability, while maximizing the end user Quality of Experience (QoE). Furthermore, insight into how different encoding parameters impact QoE can be useful to cloud gaming providers in terms of potential resource savings (e.g., if QoE remains high even while decreasing framerate to 25 fps for a certain type of game

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2016 ACM. ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

and available bandwidth, then there is no need to stream at 60 fps). We use empirical results obtained from a controlled subjective laboratory study involving 52 participants and two game types to derive analytical QoE estimation models as functions of bitrate and framerate. We further consider the impact of contextual factors including game type and player skill on QoE model specification. In addition to subjective scores, we report on objective video metrics aimed to characterize the different games used for test purposes.

The paper is organized as follows. In Section 2 we give a comparative summary of related studies that are relevant for modeling and managing QoE in the context of cloud gaming. In Section 3 we present the test methodology used to conduct subjective user studies, including the laboratory setup, participants, and test content. Section 4 gives an overview of obtained subjective scores, which are used to derive QoE estimation models based on assigned framerate and bitrate. Moreover, based on video traces collected during gameplay, we extract spatial and temporal video metrics, aimed at objectively classifying different games for the purpose of choosing an appropriate and QoE-driven video codec configuration strategy. Section 5 provides concluding remarks and directions for ongoing and future work.

2. RELATED WORK

Over the past years there have been significant research efforts in the domain of cloud gaming aimed at studying the relationships between end-user QoE and various network, service, and context factors. While many earlier studies focused on traditional online gaming have provided insight into user-level requirements in terms of factors such as perceived end-to-end latency [4], cloud gaming traffic is inherently different and thus calls for new studies to determine how certain network (e.g., latency, loss) or application-level (e.g., video encoding, content) factors map to user perceived quality metrics. In Table 1 we give a detailed overview of subjective studies that have focused on measuring and modeling QoE for cloud gaming. The table contains, for each work, the information about the platform on which the tests have been conducted, influence factors which have been tested (e.g., latency, frame rate), number of test participants in the study, measurement methodology, and identified results relevant for QoE modelling.

In terms of test platform used, numerous studies have been conducted using the GamingAnywhere platform, an open cloud gaming system that allows researchers to perform repeatable experiments and confirm reliability of their study findings [2, 7, 11, 19]. Other platforms used have included Steam [20], OnLive [7, 14, 8, 18], Ubitus [24], or other experimentally set-up platforms.

With respect to tested QoE influence factors, a large number of studies have focused on the impacts of latency and/or packet loss on user perceived quality [7, 19, 24, 15, 13, 18, 8, 17, 14, 23], while fewer studies have addressed the impact of different video encoding configurations on QoE [11, 20, 15, 2, 23]. Furthermore, while certain studies are focused on developing models for estimating actual user QoE [23], such models can be too complex (in terms of number of predictors considered) and thus not applicable in the context of application and cloud resource adaptation (i.e., when attempting to reconfigure video codec parameters on the fly) [11]. Finally, due to the high complexity of modeling gaming QoE given a wide range of game genres, end user skill levels, and end user device capabilities, challenges lie in collecting empirical ground truth data to be used for deriving accurate and applicable QoE estimation models [16].

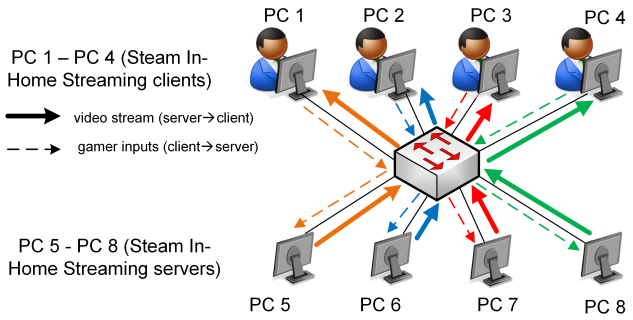


Figure 1: Laboratory testbed

As stated previously, in this paper we aim to build on and complement the results given in [11] and in [20] with the aim of contributing to further insight into video encoding adaptation strategies. We note the following key differences with respect to these studies:

1. While Hong *et al.* [11] use a crowdsourcing approach (which yields a large number of subjective ratings but without a controlled environment), we conduct tests in a controlled lab environment with all users playing on the same equipment, so as to eliminate the noise due to players playing using heterogeneous network and device conditions;

2. We test different games than those tested in [11], and look to compare with the previously published QoE modeling results. As compared with [20], we test one of the same games, but at different bitrate and framerate levels;

3. We use the Steam platform as opposed to the GA platform used in [11] so as to test commercial-grade streaming software;

4. We conduct tests at higher framerate and bitrate conditions, thus covering a different spectrum of test conditions as compared to both [11] and [20], and

5. We address the impact of user skill level on QoE, which is not explicitly addressed in [11]. Moreover, while skill level is addressed in [20], we conducted tests at different test conditions (higher bitrate and framerate levels) so as to determine deeper insight into thresholds corresponding to perceived QoE degradations.

Further discussions are given in Section 4.

3. METHODOLOGY

3.1 Subjective tests

The QoE study consisted of participants taking part in a two and a half hour long gaming session that was conducted in a laboratory environment as shown in Figure 1. Valve's Steam In-Home streaming platform was used as the cloud

Table 1: Overview of studies addressing cloud gaming QoE

Author (Year)	Cloud gaming platform	Tested QoE influence factors			Number of test participants and type of QoE study	QoE measurement methodology	Relevance for QoE modeling
		Network factors	Video factors	Context factors			
Hong et al (2015) [11]	GA	-	Frame rate, bitrate	Game genre	101; crowdsourced study	7-pt. ACR scale	Proposed gaming QoE MOS (Mean Opinion Score) model as a quadratic function of video encoding parameters (frame rate, video bitrate)
Slivar et al (2015) [20]	Steam platform	-	Frame rate, bitrate	Game genre, player skill	15; controlled lab environment	5-pt. ACR scale; Overall QoE and its features, willingness to play	Modelled QoE as a linear function of video frame rate and bitrate
Claypool et al (2014) [7]	OnLive & GA	Latency	-	Game genre, different type of end-user's device	49 (OnLive), 34 (GA); controlled lab environment	7-pt. ACR scale (OnLive); 5-pt. ACR scale (GA); Gameplay experience	Cloud-based games (regardless of the genre) are as sensitive to latency as FPS games in traditional online gaming (the most sensitive genre)
Slivar et al (2014) [19]	GA	Latency, packet loss	-	Player skill	35; controlled lab environment	5-pt. ACR scale; Overall OoE and its degradations, willingness to play	Modelled QoE as a linear function of network delay and packet loss
Wen et al (2014) [24]	Ubitus	Latency, bandwidth	-	Game genre, PC set-up, game special effects	14; controlled lab environment	5-pt. ACR scale; Video and gameplay smoothness, graphics quality	MOS of all measured QoE components are strongly corellated with network delay
Liu et al (2014) [15]	Exper. set-up	Latency, packet loss	Frame rate, bitrate	Game genre, game content (view distance, texture detail, etc.)	18 (first study), 23 (second study); controlled lab environment	5-pt. ACR scale; CMR-MOS	Proposed a content-aware model for mobile cloud gaming as a function of graphic rendering, video encoding and networking factors
Ahmadi et al (2014) [1]	-	-	-	Game genre, game content	20; controlled lab environment	5-pt. ACR scale	Proposed a game attention model for efficient bitrate allocation in cloud gaming
Beyer et al (2014) [2]	GA	-	Bitrate	-	32; controlled lab environment	GEQ, EEG	Low video quality imposed by low video bitrate has significant effect on participant's satisfaction
Jarschel et al (2013) [13]	Exper. set-up	Latency, packet loss	-	Game genre	58; controlled lab environment	5-pt. ACR scale; Overall QoE, willingness to pay	Identified key influence factors for cloud gaming QoE
Quax et al (2013) [18]	OnLive	Latency	-	Game genre	8; controlled lab environment	7-pt. Likert scale & GSR; Perceived gameplay experience, enjoyment and frustration	Latency has similar impact on QoE for the different genres in cloud gaming as in traditional online gaming
Clincy et al (2013) [8]	OnLive	Latency, packet loss	-	-	50; controlled lab environment	5-pt. ACR scale; 8 categories of QoE used to derive QoE index;	In cloud gaming, FPS players are more sensitive to network impairments then RPG players
Möller et al (2013) [17]	Exper. set-up	Latency, packet loss, bandwidth	-	Game genre, player skill	19; controlled lab environment	7-pt. ACR scale; 7 quality aspects of QoE	Complexity of activity in game scene should be considered as influencing factor on QoE
Lee et al (2012) [14]	OnLive	Latency	-	Game genre	15; controlled lab environment	fEMG	Proposed a model that can predict real time-strictness of a game based on user input rate and game dynamics
Wang et al (2009) [23]	Exper. set-up	Latency, packet loss	Frame rate, video resolution	Game genre	21 & 15; controlled lab environment	GMOS (Game Mean Opinion score)	Proposed a model for mobile cloud gaming user experience based on manipulated factors in the study

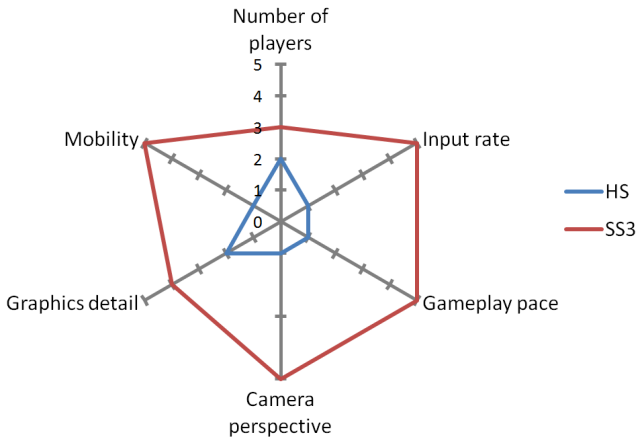


Figure 2: Comparison of game characteristics for Serious Sam 3 and Hearthstone. Points on the different axis represent different levels as described in the text.

gaming environment [1] the Steam client application was installed on all PCs in the laboratory, thus converting PC1-PC4 (Windows 7 desktops, each with Intel 3.3 Ghz i3 processor, 4GB RAM and GIGABYTE Radeon R7 250 graphic card) to Steam In-Home Streaming clients (cloud gaming clients) and PC5-PC8 (Windows 8 desktops, each with Intel 3.6 Ghz i7 processor, 8GB RAM and ASUS GT740 OC graphic card) to Steam In-Home Streaming servers (cloud gaming servers). Each of the clients had a corresponding Steam In-Home Streaming server associated, therefore four participants were able to play simultaneously during the experiments.

Two games were played in the study as follows: *Serious Sam 3 (SS3)*, representing a fast paced first person shooter game, and *Hearthstone (HS)*, a relatively slow paced card game. We illustrate the differences between these two games in Figure 2 and according to the following characterization dimensions (inspired by the categorization given in [21]): number of players, input rate, gameplay pace, camera perspective, graphics detail, and mobility of avatars. Each dimension is divided into five levels, except for camera perspective, which is divided into three levels based on [5]. The number of players is divided into five levels (from 1 to 5): single player games, two-player games, games intended for up to ten players, games intended for up to 100 players, and games for more than 100 players. In this dimension HS is placed into category 2 and SS3 into category 3. Input rate is divided based on average action per minute rate (APM) into the following categories: <10 APM, between 10 and 20 APM, between 20 and 30 APM, between 30 and 40 APM, and 50 and more APM. In this dimension HS is placed into category 1 and SS3 into category 5. The gameplay pace is specified based on the rate of the events in the game which require player reaction. In this dimension, HS is placed into category 1 as the pace is very low (usually players need to react to 1 or 2 events in 70 seconds). SS3 is placed into category 5 as the rate of events (i.e., attackers in the game) can be even multiple in one second. We opted to use these two games in our study as they represent two ends of the spectrum on many of the defined dimensions. Both games were played in HD-ready resolution (720p) with default graphics settings.

The participants in our subjective tests were 52 students enrolled at the University of Zagreb, 38 male and 14 female adults, aged between 21 and 26 (median age 23). Prior to the experiments being conducted, the participants were instructed to fill in an online questionnaire, so as to obtain relevant information about their previous overall gaming experience and gaming experience relevant to the tested games. As a result, 16 novice, 22 intermediate skilled and 14 self reported experienced players took part in the study. Since previous studies for traditional online gaming have shown that players' group composition based on previous gaming experience has an impact on perceived QoE [22], test groups were formed accordingly to investigate if this phenomenon occurs similarly in cloud gaming. The participants were organized in 13 groups with 4 players in each group, based on their reported gaming experience (skill). Each of the formed heterogeneous groups had one novice and one experienced player, while homogeneous groups consisted of 4 players with the same gaming skill level.

As stated previously, our focus in this paper is not on analysing the impact of network parameters on cloud gaming (as has been addressed in many previous studies), but rather on the investigation of the impact of video encoding parameters on QoE, with a focus on the cloud game provider perspective. Therefore, we manipulated video frame rate and bitrate, consequently controlling/influencing image quality and smoothness of gameplay. Our aim was to investigate how and to what extent these parameters affect perceived QoE for different types of games, with the ultimate goal being to use this information to derive codec configuration strategies and optimized resource allocation (from a network/service provider standpoint), while at the same time preserving high QoE. For the manipulation of video frame rate, we decided to use four levels of frame rate: 25 fps, 35 fps, 45 fps and 60 fps. In the aforementioned previous studies [11, 20], the lower end of the fps spectrum was investigated, so we opted for relatively higher values of frame rate, which coincides with the expectations of average experienced gamers regarding video frame rate. As far as video bitrate is concerned, we selected three levels for the experiments: 3 Mbps, 5 Mbps and 10 Mbps. Both framerate and bitrate were manipulated through Steam's developer console. We note that we had to limit ourselves to a certain number of test conditions, constrained by the length of subjective testing sessions. Additional test conditions would potentially lead to overly lengthy gaming sessions and possibly player fatigue. The chosen test conditions were based on our aim to complement previous studies, in the sense that we address conditions under which the impact of different bitrate/framerate combinations on QoE has not been well studied. Furthermore, prior the user study, we conducted tests to check if our testbed set-up has sufficient hardware and software capabilities necessary to support all tested games and conditions. We measured performance (frame rate) of our testbed for each tested condition and it proved sufficient for all conditions.

Considering manipulated video encoding parameters and different games, a total of 24 different test conditions were investigated during this study, with all conditions tested by each test group. To avoid bias of manipulated video pa-

[1] Steam In-Home streaming, http://store.steampowered.com/streaming/

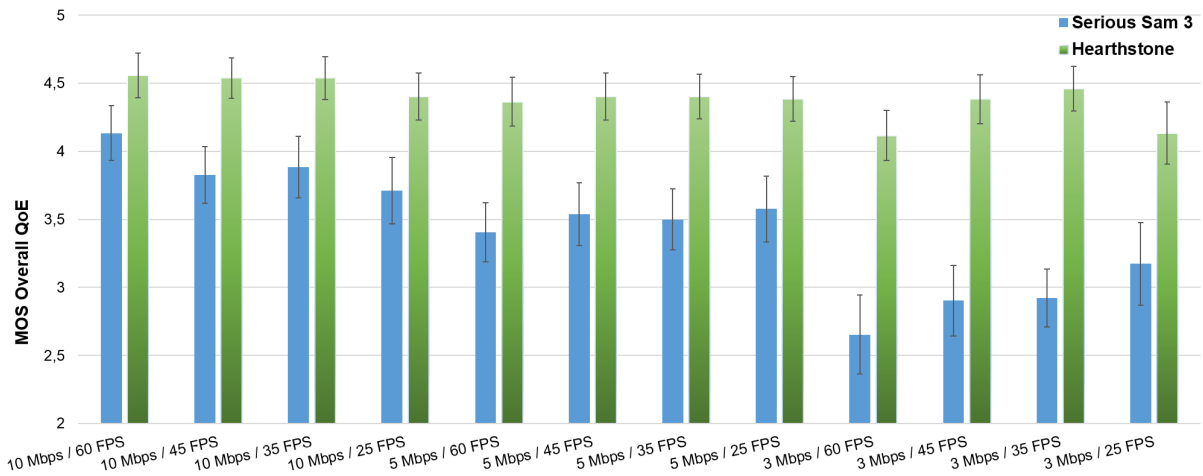


Figure 3: Subjective ratings of overall QoE (95% CI)

rameters, the sequence of test scenarios was randomized for each group. At the very beginning of the experiment, the participants were familiarized with the concept of cloud gaming and the Steam In-Home Streaming service. All the participants from each test group were seated in the same experimental room, with PCs located next to each other in one row (the participants could see each others screens and communicate with each other during experiments). Before tests started, the participants were given a short time to familiarize with game specific mechanics and the chosen map. The first 12 test scenarios involved playing one round of SS3 cooperative survival mode on a single map. During these test scenarios, it was expected from the participants to cooperate with each other to survive longer on the map. Each of these 12 test scenarios lasted on average from 2 to 5 minutes, depending on how long players from the test group survived. After finishing each test scenario, the participants were instructed to report *overall QoE*, *perceived graphics quality* and *perceived fluidity of gameplay* (on a 5-pt. ACR scale). Fluidity was explained as referring to the perception of the smoothness in the rendering of the virtual scene. Additionally, participants also reported their willingness to continue playing under the given test conditions for the current test scenario (yes/no). We also recorded the survival time for each player. While participants were filling in questionnaires, the test administrator changed the video encoding parameters by running scripts on the player's PCs. The second half of the experiment involved playing HS. HS is a digital card game that consists of turn-based matches between two players. For that reason, an opponent from the group was assigned to each player by the test administrator. In the case of HS, each test scenario lasted 3 minutes, after which the participants filled in a questionnaire and continue playing the ongoing match. The entire gaming session (with a 10-minute break allotted in the middle) lasted approximately two and a half hours, depending on the group's performance during the SS3 test scenarios. We note that that potential order effects may have occurred during experiments due to the experimental design (order of games).

3.2 Video characterisation

To be able to empirically quantify the differences between the two tested games and relate them to QoE, we analysed both temporal and spatial characteristics of their video streams. The first set of metrics is extracted according to ITU-T recommendation P.910 (4/2008): Spatial perceptual information (SI) and Temporal perceptual information (TI) [12]. The second set of metrics is extracted according to [3]: Percentage of Forward/backward or Intra-coded Macroblocks (PFIM) for the temporal aspect of the video (motion in subsequent images), and Intra-coded Block Size (IBS) for the spatial aspect of video (scene complexity).

SI is derived based on the Sobel filter. Each video frame (luminance plane) at time n (F_n) is first filtered with the Sobel filter [$Sobel(F_n)$]. The standard deviation over the pixels (std_{space}) in each Sobel-filtered frame is then computed. This operation is repeated for each frame in the video sequence and results in a time series of spatial information of the scene. The maximum value in the time series (max_{time}) is chosen to represent the spatial information content of the scene. This process can be represented in equation form as:

$$SI = max_{time}\{std_{space}[Sobel(F_n)]\} \quad (1)$$

More details in the frame will result in higher values of SI.

TI is based upon the motion difference feature, $M_n(i,j)$, which is the difference between the pixel values (of the luminance plane) at the same location in space but at successive times or frames. $M_n(i,j)$ as a function of time (n) is defined as:

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j) \quad (2)$$

$F_n(i,j)$ is the luminance value of the pixel at the ith row and jth column of nth frame in time. The measure of temporal information (TI) is computed as the maximum over time (max_{time}) of the standard deviation over space (std_{space}) of $M_n(i,j)$ over all i and j.

$$TI = max_{time}\{std_{space}[M_n(i,j)]\} \quad (3)$$

More motion in adjacent frames will result in higher values of TI. For scenes that contain scene cuts, two values may be given: one where the scene cut is included in the temporal information measure, and one where it is excluded from the measurement (in our case no scene cuts were not present and normal gameplay was recorded). TI and SI metrics have

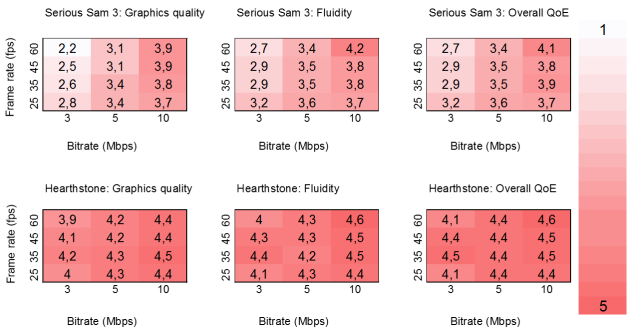


Figure 4: Aggregated subjective ratings for each game under different video configurations

been extracted through predefined Matlab scripts (authored by Savvas Argyropoulos).

The logic behind PFIM and IBS metrics is as follows: A video with visual changes from frame to frame will have these changes encoded (either by neighbouring blocks or independently of other blocks), while video without visual changes can skip much of the encoding (PFIM) and if the scene is simple, there is not much information to be encoded. As a result, the intra-coded block size will be small. If the scene is complicated, the IBS will be large to contain all the information. PFIM and IBS metrics were extracted using python scripts created by Mark Claypool [3].

4. RESULTS

4.1 Subjective results

Figure 3 shows the average subjective ratings of overall QoE for SS3 and HS across all test conditions. First of all, it can be observed that there is a significant difference between overall QoE for both games: HS has on average higher scores of overall QoE for all test conditions in comparison with SS3, with the average QoE score never going below 4.0 for any given test scenario. Furthermore, we observe that neither lowering video frame rate nor video bitrate had a significant impact on perceived QoE during HS gaming sessions. We can assume that during our experiments, the manipulated frame rate and bitrate values were high enough that the participants did not perceive QoE degradations for HS. On the other hand, average QoE scores for SS3 are significantly lower than HS QoE scores for each of the test scenarios, with only one test scenario (the case with 60 fps and 10 Mbps) averaging more than 4.0 score. Moreover, it can be noticed that manipulation of video encoding parameters significantly affects perceived QoE for SS3 gaming sessions: when bitrate values are high enough (10 Mbps), lowering frame rate leads to degradations of QoE. SS3 is a representative fast paced first person shooter game, thus degradations of fluidity (smoothness of gameplay), introduced by lowering frame rate, have a higher impact when the bitrate is high enough to support transmission of high quality video. However, for low bitrate levels (3 Mbps), average scores of perceived QoE are ascending with reductions of frame rate (down to 25 fps). This can be attributed to the fact that 3 Mbps bitrate is not high enough to preserve *good enough* video quality, so even though fluidity is very important for fast paced games, the participants do not tolerate

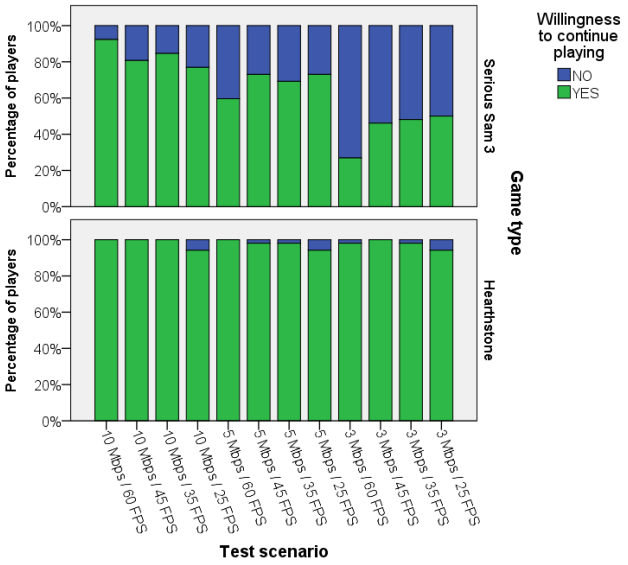


Figure 5: Willingness to continue playing under different test conditions for both tested games

low graphics quality and thus prefer an increase in graphics quality at the expense of lowering the fluidity of gameplay for these scenarios. Given these results, we see that *different encoding strategies may be employed for different games to maintain high player QoE.*

Besides collecting data about overall QoE scores we collected data about user perceived fluidity and graphics quality (such measures have also been reported in related work [11]). A heatmap overview of collected data (Figure 4) shows the mean scores for overall QoE, graphics quality and fluidity. There is a very large correlation between the measured metrics indicating that players form an opinion about the test scenario and score the different dimensions based on this opinion. It can be noticed that the HS MOS score for overall QoE and its features (fluidity, graphics quality) are on average much higher and are prone to minor changes due to manipulation of video parameters in comparison with MOS scores for SS3. This further supports the claim that the majority of players do not easily perceive QoE degradations while playing a slow paced game such as HS at a higher spectrum of test conditions.

In addition to differences in aggregated scores, there is also a large discrepancy in the number of test scenario where the participants were not willing to continue playing under current test conditions between tested games, as shown in Figure 5: for SS3, there were 218 occurrences (from 624 overall) when players stated they would not continue playing under the given conditions, while for HS there were only 13 cases (from 624 overall) when players stated they would quit playing. It can be observed that for 3 Mbps and 60 fps 73.1% of players were not eager to continue playing SS3, while for HS under the same test conditions only 1.9% players wanted to quit playing. Additionally, we observe that at a bitrate of 3 Mbps, a decrease in frame rate actually results in an increase in the percentage of players reporting they would continue playing SS3, whereas for HS the same manipulation of frame rate does not result with such an increase in the percentage of players willing to continue playing. This

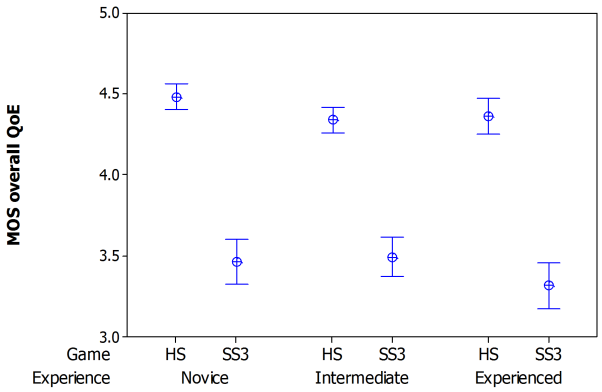


Figure 6: Subjective ratings of overall QoE (95% CI) for SS3 and HS grouped by skill

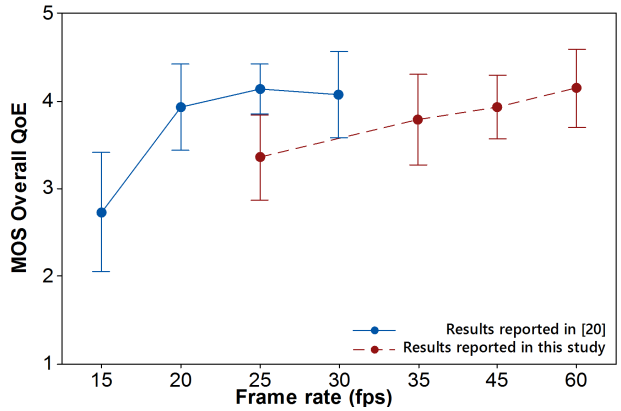


Figure 8: MOS scores for different frame rates at a fixed bitrate of 10 Mbps for SS3

furthermore confirms the need for deriving video encoding adaptation strategies for different types of games when aiming to optimize end user QoE. Given the length of the user study (2.5 hrs), we also tested whether there was an impact on user fatigue and trends in user ratings from beginning to the end of the test session. Thus, we compared the overall user ratings for conditions tested early in the test session and those tested late in the session (note test ordering was randomized). No clear differences or trends were observed.

4.1.1 User parameters

Another goal of our study was to examine the impact of user parameters on QoE, primarily in terms of player's previous gaming experience. Overall QoE ratings for SS3 and HS grouped by player experience are shown in Figure 6. While experienced players gave on average slightly lower QoE scores in accordance with the introduced QoS degradations in comparison with less experienced players, the confidence intervals are overlapping so no clear statistical distinction can be made. On the other hand, reported overall QoE scores for intermediate players varies. In case of SS3, intermediate players have on average slightly higher scores than novice players, but in case of HS we observe that their scores are slightly lower when compared with experienced players. We note that the ratings in Figure 6 represent aggregate scores. When analysed on a per test scenario basis (Figure 7), experienced players tended to give lower scores for lower quality scenarios, and higher scores for higher quality scenarios, as opposed to novice players. This may be attributed to the hypothesis that novice players are generally less sensitive to different quality variations (this was also visible when considering the distributions of scores across all scenarios per skill level).

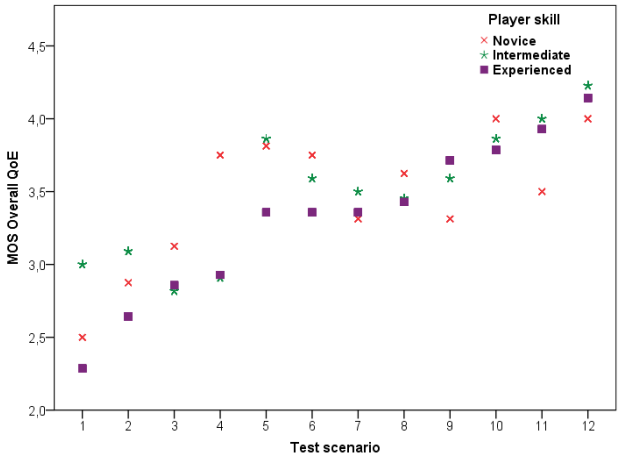


Figure 7: Mean ratings of overall QoE per skill level for different test scenario (scenarios arranged according to ascending mean QoE

We further compare the overall QoE scores of experienced players considered in this study with the results from the study reported in [20] which considered only experienced players. A comparison of scores is given in Figure 8, showing average QoE ratings for different fps conditions at a set bitrate of 10 Mbps. We note differences between methodologies (e.g., different tested fps levels) and context (single-player mode was used in [20] while multiplayer mode was used in this study). In [20], the range of tested frame rates was 15 – 30 fps, while in this study we tested 25 – 60 fps. Consequently, the "best" scenarios observed by players (in terms of fps) in [20] were 25 and 30 fps, while in this study 25 fps was the "worst" tested value. Interestingly, the same test condition (25 fps, 10 Mbps) was rated quite differently in these two studies, which can likely be attributed to the choice of tested stimuli and player tendency to compare conditions relative to one another. This raises several important questions regarding: use of rating scales and comparison of results, specifications of different contexts (i.e., single vs group play), and the implications of ranges of tested system parameters (in this case frame rate). We note that recent efforts have aimed at addressing the challenge of standardizing test methodologies for gaming QoE [16].

Referring again to the results of this study, we illustrate the extent of the degradations introduced in Figure 9. Three areas of degradations can be identified (as presented in [10] regarding generic relationships between QoE and QoS): (1) no distortion perceived, (2) user disturbed and (3) user gives up. During this study and in the case of SS3, most of the player scores are located in the "user disturbed" area. For the case of HS, the tested degradations did not have a signif-

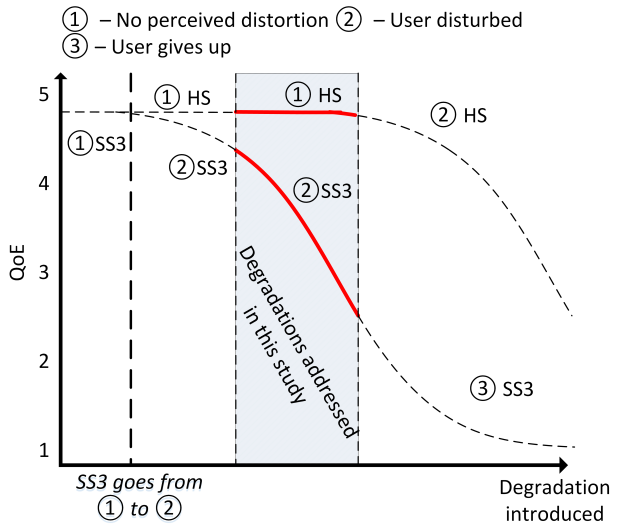


Figure 9: Impact of degradations on QoE

icant impact on perceived QoE, and thus player scores are portrayed in the "no distortion area". We note that Figure 9 is meant to only illustrate qualitative trends, and that future studies are needed to assign concrete values to all points along the axis.

4.1.2 Context parameters

We additionally inspected the impact on QoE of the players' social context. Social context is represented by players' group composition based on previous player's gaming experience. Out of 13 groups that participated in our study, 2 of them were composed of only experienced players, 4 groups were composed of intermediate skilled players and 2 groups included only novice players. The remaining 5 groups were heterogeneous groups with regards to previous player's experience, and each of these mixed groups consisted of at least one novice and one experienced player. Figure 10 displays average scores of overall QoE for SS3 based on group composition. The distinction of QoE scores between homogeneous and mixed groups is minor across all experience

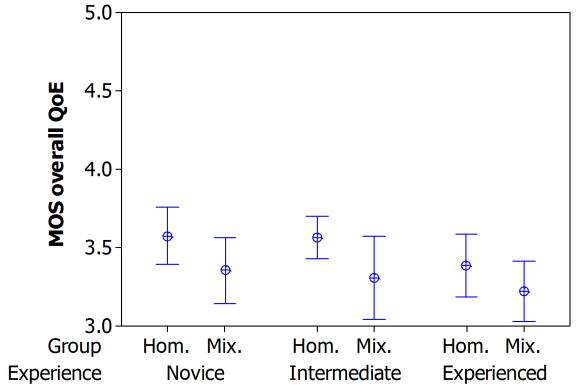


Figure 10: The impact of group composition on QoE for SS3 (avg. values and 95% CI)

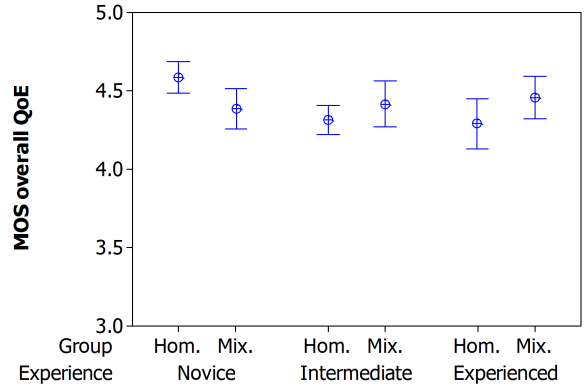
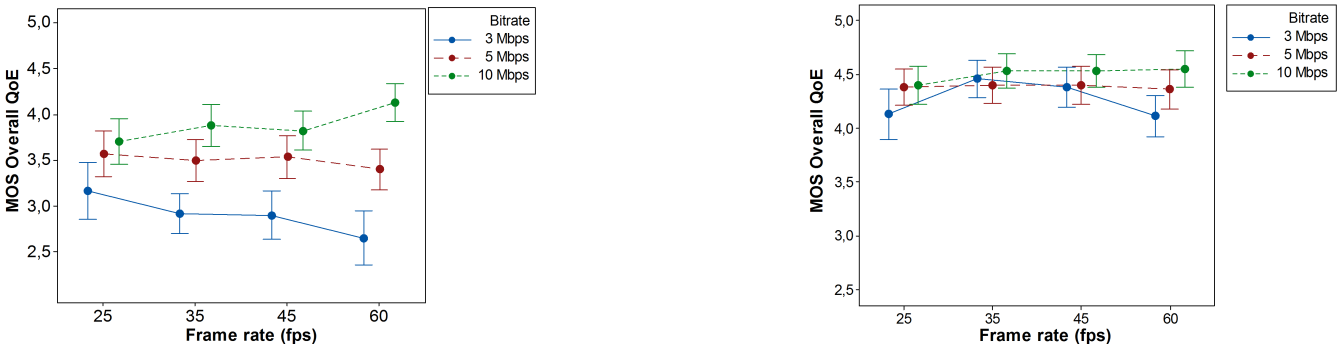


Figure 11: The impact of group composition on QoE for HS (avg. values and 95% CI)

levels, although a slight decrease of perceived QoE can be observed when playing in mixed groups for all levels. This differs from findings in [22], where only experienced players reported lower QoS scores in mixed groups, while for novice and intermediate player playing in mixed groups improved QoE, due to playing with experienced players which yielded better game performance results for less experienced players. However, group composition has a different impact on QoE for HS (Figure 11). While novice players report lower QoE scores in mixed groups, perceived QoE of intermediate and experienced players is slightly improved while playing in the same group composition. This can be attributed to the nature of the tested games. Whereas SS3 was played cooperatively in our study, HS is a game where two players play against each other and only one of the players wins. This sometimes results with imbalanced game sessions where novice and experienced players are paired against each other, and in these types of game situations more experienced players generally win with ease, making gaming sessions more enjoyable for winners, as also reported in [6]. However, these results are only indicative and the number of test subjects is rather low for each category leading to very broad and overlapping confidence intervals. Further testing of this parameter is needed for it to be quantified and incorporated into future QoE models.

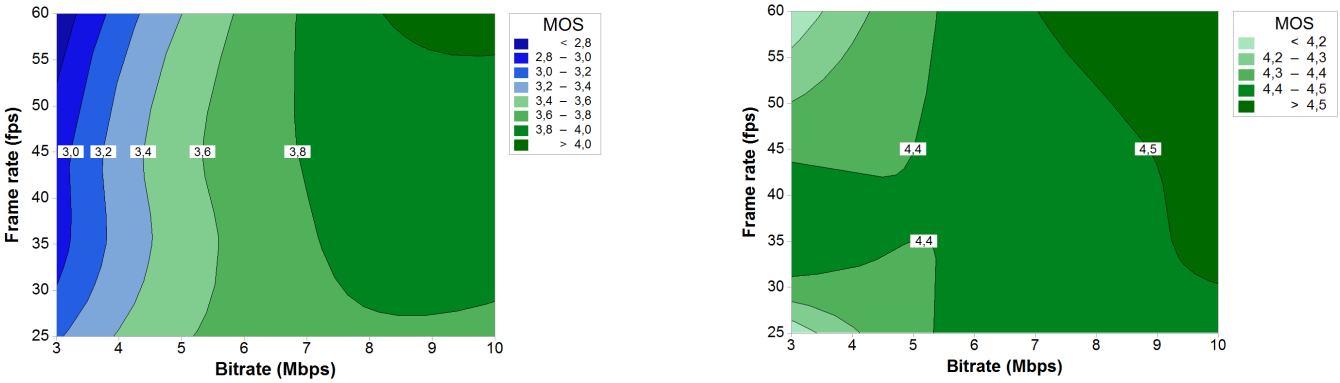
4.1.3 System parameters

The impact of frame rate on subjective ratings of overall QoE under fixed bitrate for both games is shown in Figure 12. The graph shows more clearly what was stated previously with regards to the impact of video encoding parameters on perceived QoE for SS3: players notice degradations of QoE due to reduced frame rate for high bitrate levels as a result of gameplay fluidity degradations. However, for low bitrate levels (especially 3 Mbps), a decrease of frame rate leads to a significant increase of graphics quality, which impacts players more than degradations of gameplay fluidity. On the contrary, for HS, players perceive QoE impairments (induced by manipulations of video encoding parameters) to a far less extent, which leads to the conclusion that different encoding configuration strategies can be employed for different types of games.



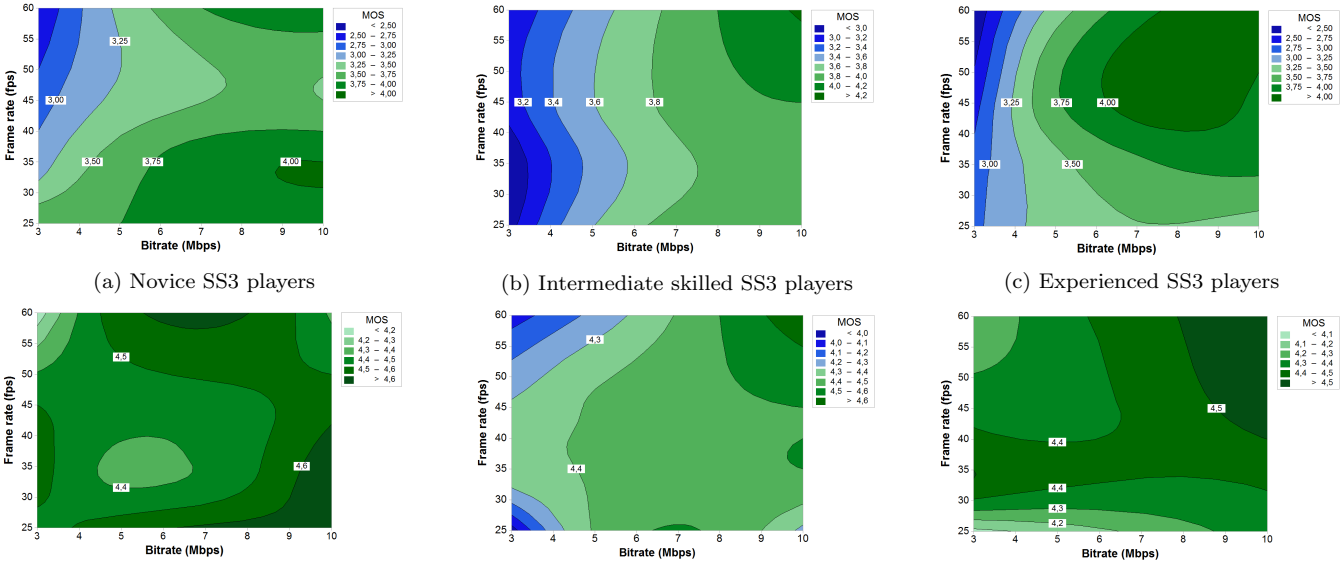
(a) Serious Sam 3 (b) Hearthstone

Figure 12: Impact of video parameters on overall MOS scores for SS3 and HS



(a) Serious Sam 3 (b) Hearthstone

Figure 13: Illustrated QoE models for Serious Sam 3 and Hearthstone



(a) Novice SS3 players (b) Intermediate skilled SS3 players (c) Experienced SS3 players

(d) Novice HS players (e) Intermediate skilled HS players (f) Experienced HS players

Figure 14: Graphical representation of QoE models for Serious Sam 3 and Hearthstone depending on player skill

Table 2: The QoE models for tested games

	Serious Sam 3				Hearthstone			
	All	Novice	Intermediate	Experienced	All	Novice	Intermediate	Experienced
framerate, $\alpha_{g,1}$	-0.028	0.199	0.466	0.541	0.034	0.072	0.107	-0.025
bitrate (Mbps), $\alpha_{g,2}$	0.404	-0.022	-0.028	-0.046	0.060	-0.003	-0.010	0.002
I($framerate^2$), $\alpha_{g,3}$	6.391E-04	-0.096	-0.009	0.019	0.060	0.014	0.039	0.049
I($bitrate^2$), $\alpha_{g,4}$	-0.031	0.001	7.701E-05	-0.001	-0.004	-2.168E-04	-0.001	-0.001
framerate:bitrate, $\alpha_{g,5}$	0.003	0.005	0.001	0.005	0.001	1.572E-04	0.002	0.001
Constant, $\alpha_{g,6}$	2.611	4.902	1.897	1.116	3,473	4.065	3.155	3.296
R^2	0.986	0.915	0.969	0.977	0.782	0.496	0.773	0.763

4.2 QoE estimation models

In this section we report on obtained QoE estimation models derived from empirical data. We note that, as stated by Hong et al. [11], such models as those proposed here are not meant to provide overall accurate QoE estimations (as has been the target of more complex models with a larger number of predictor variables, including various context factors, latency, etc.), but rather to provide input to the cloud service provider in terms of codec (re)configuration in light of available bandwidth.

To model QoE as a function of video encoding parameters, we tried several different ways to fit the data by using different types of linear and non linear models. It should be noted that we consider our data as interval data and not ordinal (i.e., we consider that the intervals between points on the rating scales are equal). Based on our collected data and by analysing accuracy of fit for these models, we model the MOS scores as a quadratic function of manipulated video encoding parameters (as also proposed in [11]):

$$MOS(g,f,b) = \alpha_{g,1}f+\alpha_{g,2}b+\alpha_{g,3}f^2+\alpha_{g,4}b^2+\alpha_{g,5}fb+\alpha_{g,6}, \quad (4)$$

where $\alpha_{g,1}$ - $\alpha_{g,6}$ are game-specific model parameters, b is video bitrate and f is video frame rate. The values of model parameters for tested games are summarized in Table 2, together with related R-squared values (the coefficient of determination indicating how well data fits a QoE model). For QoE models where all players, regardless of experience, are considered, it can be seen that the derived QoE model for SS3 has a better fit considering collected data ($R^2 = 0.986$) then the QoE model for HS ($R^2 = 0.782$). These QoE models are visualized in Figure 13, whereby the QoE model for SS3 is visually similar to the QoE model for Call of Duty reported in [11].

In addition to modelling QoE as a function of bitrate and frame rate, we have also taken into account players' experience and modelled QoE separately for different player experience levels. We illustrate obtained models in Figure 14. As previously stated, experienced players are expected to be more aware of game impairments due to QoS degradations, and in previous studies have been shown to rate perceived QoE with lower scores than novice players. In case of SS3, novice players' QoE scores are not consistent with video quality deteriorations through test scenarios, e.g. for fixed 10 Mbps bitrate their perceived QoE is higher at lower frame rates, which conflicts with QoE scores from other skill groups. Figures 15 and 16 show the accuracy of the acquired prediction models (considering all player types) for QoE for SS3 and HS. If we use the QoE model designed without considering player experience to estimate overall QoE for different skilled players, we see that there are wide discrepancies between estimated and reported values of QoE, resulting in lower QoE estimation accuracy. We therefore propose to consider player skill as one of the inputs for QoE modelling. While in this work we propose different QoE models for different skill levels, future work will address the potential of incorporating skill level into a more generalized QoE estimation model.

4.3 Characterization of test content based on objective video metrics

Finally, we report on measured objective video metrics. We collected gameplay video traces at a video encoding frame rate of 30 fps for all three tested bitrate levels, which resulted in 156 video samples per game. Figures 17 and 18 show plots of scores for HS and SS3 for temporal and spatial metrics. In Figure 17 we plot temporal (TI) and spatial (SI) metrics against each other. It can be seen that both games

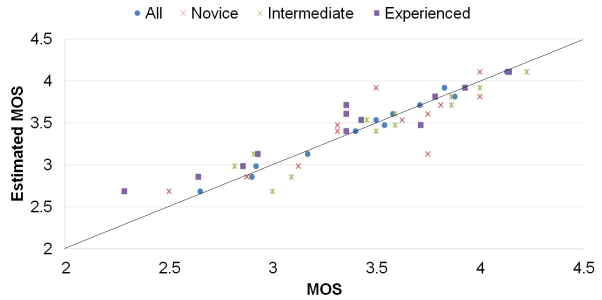


Figure 15: Accuracy of estimated QoE ratings vs subjective QoE ratings for Serious Sam 3

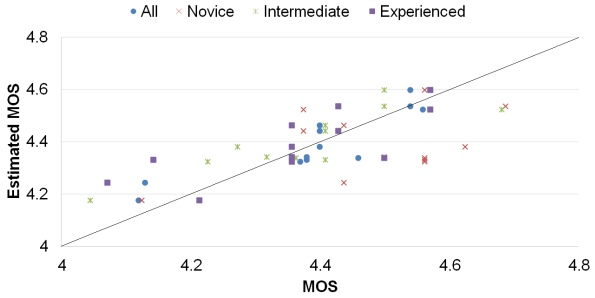


Figure 16: Accuracy of estimated QoE ratings vs subjective QoE ratings for Hearthstone

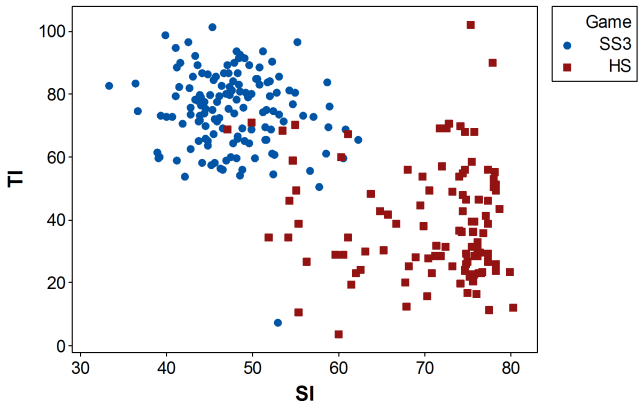


Figure 17: Scores for TI and SI metrics for SS3 and HS

in this subfigure show clear clustering behaviour, although it is somewhat more prominent for SS3 which is clustered in the upper left portion of the graph (higher temporal dynamics and lower image details), while HS is concentrated mostly in the lower right portion of the graph (higher image details and lower temporal dynamics). It should be noted that on the temporal axis, HS is much more "spread" in values meaning that in some videos, highly dynamic actions occur due to some cards having advanced animations (e.g., Twisting Nether card). It is evident that SS3 stream is much more dynamic as scores on temporal metrics are, in general, double the HS scores. The same trends can be observed when two metrics from the other metric set (PFIM and IBS) are plotted against each other in Figure 18. For these metrics there is less spread for the temporal component for HS, while slightly more for SS3. These results indicate that it is possible to empirically quantify the differences between video streams of separate games. This information coupled with other information regarding the game can and provide a basis for future automatic game classification which can be used for selecting optimal adaptation strategies for cloud gaming.

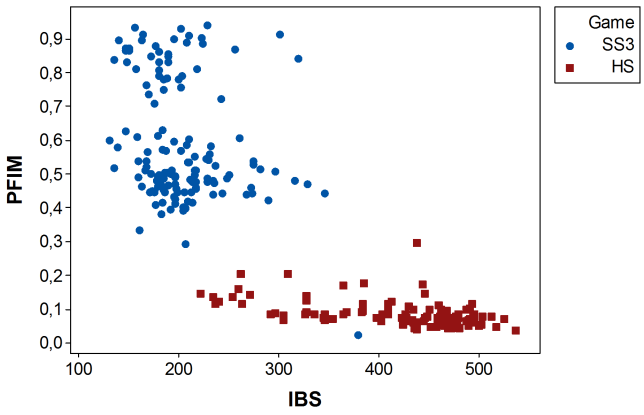


Figure 18: Scores for PFIM and IB metrics for SS3 and HSS

5. CONCLUSIONS

In this work we have presented a subjective study involving 52 players playing two games delivered via cloud gaming. We inspected the relation between system parameters (i.e., bitrate and frame rate) to perceived fluidity, graphics quality and overall QoE. We also investigated the impact of user parameters (i.e., user game experience) and context parameters (i.e., social context) on the QoE scores of both games. Based on obtained results and also inspired by previous work published in [11], we derived QoE models for these two games and show that they are significantly different. Hence, we conclude that the game type tested clearly needs to be taken into account when evaluating the QoE of cloud games. The results indicate that there is no linear relationship between frame rate and QoE – in some cases it is better to deliver lower frame rate and increase graphics quality. Also, we confirmed that there is significant impact of players' previous gaming experience and incorporated this fact into our models, while we concluded that for social context more research is needed in order to be able to numerically quantify its impact. Future studies will look to address the limitations of the study reported in this paper, including the sample test population (which should in the future be extended to better represent a wider gamer population), and test methodology (potential impact of ordering effects).

Further, our ongoing work is aimed at further deriving QoE-driven video encoding adaptation strategies for different available network conditions. We are also looking to study whether or not different adaptation strategies can be mapped to classes of games (delivered via the cloud gaming paradigm) grouped according to objective spatial and temporal video metrics, and potentially additional relevant context data (e.g., end device capabilities, player skill). Preliminary results show that measured video characteristics of the two tested games indicate that examining objective video characteristics can be a basis for game classification. Such a classification could then be used for determining optimal adaptation strategies for classes of games, which could in the future automate the process of deciding on the best encoding adaptation strategy for a particular game, alleviating the need to conduct subjective studies for additionally considered (or newly emerging) games.

We would also like to extend the reported QoE models by taking into consideration additional context, such as player skill and social context, and incorporating system parameters such as latency and packet loss, which could improve model accuracy. Finally, we would like to test the impact of different encoding adaptation strategies on client device battery consumption, as additional potential motivation to adapt codec configuration parameters.

6. ACKNOWLEDGEMENTS

This work has been supported in part by the Croatian Science Foundation under the project UIP-2014-09-5605 (Q-MANIC) and the project "Information and communication technology for generic and energy-efficient communication solutions with application in e-/m-health (ICTGEN))" co-financed by the EU from the European Regional Development Fund.

7. REFERENCES

[1] H. Ahmadi, S. Z. Tootaghaj, M. R. Hashemi, and S. Shirmohammadi. A game attention model for

efficient bit rate allocation in cloud gaming. *Multimedia Systems*, 20(5):485–501, 2014.

[2] J. Beyer, R. Varbelow, J.-N. Antons, and S. Möller. Using electroencephalography and subjective self-assessment to measure the influence of quality variations in cloud gaming. In *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, pages 1–6, May 2015.

[3] M. Claypool. Motion and scene complexity for streaming video games. In *Proceedings of the 4th International Conference on Foundations of Digital Games*, pages 34–41. ACM, 2009.

[4] M. Claypool and K. Claypool. Latency and player actions in online games. *Commun. ACM*, 49(11):40–45, Nov. 2006.

[5] M. Claypool and K. Claypool. Perspectives, frame rates and resolutions: it's all in the game. In *Proceedings of the 4th International Conference on Foundations of Digital Games*, pages 42–49. ACM, 2009.

[6] M. Claypool, J. Decelle, G. Hall, and L. O'Donnell. Surrender at 20? Matchmaking in League of Legends. In *Games Entertainment Media Conference (GEM), 2015 IEEE*, pages 1–4. IEEE, 2015.

[7] M. Claypool and D. Finkel. The effects of latency on player performance in cloud-based games. In *Proc. of 13th Annual IEEE/ACM NetGames*, pages 1–6, 2014.

[8] V. Clincy and B. Wilgor. Subjective evaluation of latency and packet loss in a cloud-based game. In *Information Technology: New Generations (ITNG), 2013 Tenth International Conference on*, pages 473–476. IEEE, 2013.

[9] M. Dick, O. Wellnitz, and L. Wolf. Analysis of factors affecting players' performance and perception in multiplayer games. In *Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games*, NetGames '05, pages 1–7, 2005.

[10] M. Fiedler, T. Hossfeld, and P. Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *Network, IEEE*, 24(2):36–41, 2010.

[11] H. Hong, C. Hsu, T. Tsai, C. Huang, K. Chen, and C. Hsu. Enabling adaptive cloud gaming in an open-source cloud gaming platform. *IEEE Transactions on Circuits and Systems for Video Technology*, PP(99):1–14, 2015.

[12] P. ITU-T Recommendation. Subjective video quality assessment methods for multimedia applications. 1999.

[13] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hoßfeld. Gaming in the clouds: QoE and the users' perspective. *Mathematical and Computer Modelling*, 57(11):2883–2894, 2013.

[14] Y.-T. Lee, K.-T. Chen, H.-I. Su, and C.-L. Lei. Are all games equally cloud-gaming-friendly? An electromyographic approach. In *Network and Systems Support for Games (NetGames), 2012 11th Annual Workshop on*, pages 1–6. IEEE, 2012.

[15] Y. Liu, S. Wang, and S. Dey. Content-aware modeling and enhancing user experience in cloud mobile rendering and streaming. *IEEE J. Emerg. Sel. Topics Circuits Syst.*, 4(1):43–56, 2014.

[16] S. Moller, J.-N. Antons, J. Beyer, S. Egger, E. Nunez Castellar, L. Skorin-Kapov, and M. Suznjevic. Towards a new ITU-T recommendation for subjective methods evaluating gaming QoE. In *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, pages 1–6. IEEE, 2015.

[17] S. Möller, D. Pommer, J. Beyer, and J. Rake-Revelant. Factors Influencing Gaming QoE: Lessons Learned from the Evaluation of Cloud Gaming Services. In *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS 2013)*, pages 1–5, Sept. 2013.

[18] P. Quax, A. Beznosyk, W. Vanmontfort, R. Marx, and W. Lamotte. An evaluation of the impact of game genre on user experience in cloud gaming. In *Games Innovation Conference (IGIC), 2013 IEEE International*, pages 216–221. IEEE, 2013.

[19] I. Slivar, M. Suznjevic, and L. Skorin-Kapov. Empirical QoE Study of In-Home Streaming of Online Games. In *Proceedings of the 14th Annual Workshop on Network and Systems Support for Games, NetGames*, Nagoya, Japan, Dec. 2014.

[20] I. Slivar, M. Suznjevic, and L. Skorin-Kapov. The impact of video encoding parameters and game type on QoE for cloud gaming: A case study using the Steam platform. In *Quality of Multimedia Experience (QoMEX), 2015 7th Intl. Workshop on*, pages 1–6, May 2015.

[21] M. Suznjevic, O. Dobrijevic, and M. Matijasevic. MMORPG player actions: Network performance, session patterns and latency requirements analysis. *Multimedia Tools and Applications*, 45(1-3):191–241, 2009.

[22] M. Suznjevic, L. Skorin-Kapov, and M. Matijasevic. Impact of User, System, and Context factors on Gaming QoE: a Case Study Involving MMORPGs. In *Proceedings of the 12th ACM SIGCOMM Workshop on Network and System Support for Games*, 2013.

[23] S. Wang and S. Dey. Modeling and characterizing user experience in a cloud server based mobile gaming approach. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1–7. IEEE, 2009.

[24] Z.-Y. Wen and H.-F. Hsiao. QoE-driven performance analysis of cloud gaming services. In *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*, pages 1–6. IEEE, 2014.