Subjective Quality Assessment for Cloud Gaming

Abstract—Quality of experience (QoE) is widely used to evaluate the user perception of network applications. In recent years, the popularity of gaming applications has increased exponentially, especially, cloud gaming. QoE of interactive cloud gaming heavily relies on the network Quality of Service (QoS) parameters such as delay, packet loss ratio (PLR) and jitter. In this paper, we study the effect of the different network parameters on the QoE of cloud gaming. Using subjective testing we evaluated the effect of network degradation on the user experience. We studied three different games, selected based on genre, popularity and content complexity, and tested them in a controlled network environment to study how different games are affected by the different network conditions.

We captured user ratings on an Absolute Category Rating (ACR) for video QoE, game QoE and overall QoE of the Game-Under-Test (GuT) experienced by the user at the set network conditions. We report our findings as Mean Opinion Score (MOS) and show that, as expected, the games with a high level of content complexity are more badly affected by network impairments than the game with a low level of content complexity. We also highlight that the QoE of interactive cloud applications relies more on the game playability than video quality of the game. These findings can support service providers in enhancing the end-user QoE of cloud gaming applications to retain user subscriptions and avoid customer churn.

Index Terms—Quality of Experience (QoE), Quality of Service (QoS), Packet Loss Ratios (PLR), Cloud gaming, interactive gaming, Video QoE, immersion, Gaming influence factors.

I. Introduction

Over the last couple of decades, rapid growth in communication technology has brought forward new services such as video streaming and cloud gaming over the internet. Due to the increased customer interest, big tech-giants such as Microsoft and Google have introduced their cloud gaming platforms.1 This is a new form of gaming and is sometimes referred to as Gaming on Demand. The idea of cloud gaming is to play on a remote server games, which are then streamed over the internet to the user’s device. The user’s device does not need the hardware capability to run the game, since all the rendering of the game will be done at the server side, saving the user the cost of getting a high-end GPU powered device [2].

Real-time applications like cloud gaming are quite demanding in terms of QoS parameters, requiring high bandwidth and low latency and PLR conditions [3]. Service providers are aware of the potential growth of the new services and are keen to keep their existing customers happy as well as gain new customers. In order to do that they have to provide competitive service level agreement to ensure satisfactory Quality of Experience (QoE) for cloud gaming applications.

QoE assessment of cloud gaming application is a trending research field. With a very limited literature on the subject, the field and the tools employed for QoE evaluation of cloud gaming applications are far from maturity. There are accounts of subjective testing being used to evaluate the effect of cloud gaming influencing factors, such video-encoder, frame rate and QoS parameters [4]. However, since game development is becoming more sophisticated, it is vital to keep up the QoE evaluation targeted for more complex and high-end games.

In this paper, we used subjective testing to evaluate the effect of various network conditions on the video quality, game playability and overall experience of the user for three games with different characteristics in terms of genre, content complexity and pace.

The results provide indications on which factor (video quality or the game-playability) contribute more towards the overall QoE of the games.

II. Related Work

Improvement in the cloud infrastructure and increased interest by the users have propelled the academia and industry to address QoE evaluation of cloud gaming applications. Authors in [2] describe the infrastructure of cloud gaming. They discuss how powerful machines in the cloud can host the game and stream it to the client’s machine over the network. The network transfers the video traffic from the cloud to the user and the user’s input to the cloud. This makes the network parameters an important influencing factor in QoE evaluation of such service.

The two main factors impacting the overall QoE of cloud gaming are the video quality and the quality of interactivity of the user with the game. Studies in [5, 6] show that video QoE tends to degrade significantly with deteriorating network performance. They reported how increasing levels of PLR can result in lower QoE for passive gaming applications.

Claypool et al [7] referred the gaming client as a thin client and studied detailed network performance when running different games on the onLive platform. Their work was extended by the authors in [8] that studied the effect of various network conditions on the objective quality metrics of two different cloud gaming platforms. They concluded that network degradation affect video and game quality on both platforms, but the onLive platform performs better than StreamMyGame (SMG) at degrading network conditions. They used frame rate as a measure of game quality.

In addition, the impact of the network parameters on the cloud gaming scenario was also reported in [9]. They replicated the cloud gaming scenario using PlayStation 3 console as a cloud machine and studied the effect of different network scenarios on the QoE of the game. They reported that at PLR>1%, the QoE perceived is unsatisfactory. Moreover,
a similar study [4] demonstrated the effect of the network conditions (frame loss, jitter, and latency) hosted on NVIDIA GeForce Now. They used subjective testing to evaluate QoE and reported that the games perform worst on increasing delay but shows resilience to the packet loss. In this paper, we use Parsec as a game hosting platform and study how the different network settings impact the video and game-playability of the user. As far as we know there are no other studies available on Parsec. We use subjective testing to evaluate the QoE and report which factor (video or game-playability) is more critical in overall QoE evaluation of the cloud gaming.

III. GAMES UNDER TEST

Three different games were selected on the basis of genre, pace, content complexity and the popularity on gaming platform. The selection was made in accordance to ITU-T recommendation G.1032 [10]. As there is very limited literature on the classification of the games [15] we tested various games and evaluated them on the basis of the aforementioned influencing factors. The games with most distinctive difference in content complexity and pace were selected. This is because according to [10] these influencing factors are critical when studying the effect of QoS parameters, e.g. delay. Table I shows a set of influencing factors against which the games were selected.

<table>
<thead>
<tr>
<th>Game</th>
<th>Genre</th>
<th>Pace</th>
<th>Learning difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grand Theft Auto V</td>
<td>Action-adventure</td>
<td>Fast</td>
<td>Easy</td>
</tr>
<tr>
<td>Counter Strike Global Offensive (CSGO)</td>
<td>First Person Shooting (FPS)</td>
<td>Medium</td>
<td>Easy</td>
</tr>
<tr>
<td>FIFA 2019</td>
<td>Sports (FPS)</td>
<td>Slow</td>
<td>Medium</td>
</tr>
</tbody>
</table>

The selected games represent a broader range of games of each genre. As there are no set guidelines to define the pace of the game, we evaluated the pace of the game on the basis of response time required by the user and the temporal complexity (explained later). Additionally, the learning difficulty of the game can also be an influencing factor [10]. This was assessed on the basis of the number of different moves and behaviours a user needs to learn to play the game. GTA V has multiple modes, but in this paper, we only considered the driving scenario. Whereas, generally for FIFA and CSGO a standard scenario includes most of the aspects of the gameplay.

A. Spatial and Temporal Information of GuT

In addition to influencing factors in Table I, the content complexity was another critical game selection criterion. This is because it is reported multiple times as an influencing factor of video streaming applications and passive gaming applications [11]. Since we want to report Video QoE of cloud gaming as one our finding, we selected games with different content complexity.

To evaluate the content complexity of the different games, we calculated spatial information (SI) and temporal Information (TI) of the recorded scenario of the game play using ITU-T recommendation P.910 [12]. We calculated SI and TI for 5 different scenarios of each games consisting of the gameplay that would be evaluated by subjective testing. The SI is calculated for each frame on the basis of the standard deviation of the luminosity plane of that frame. The SI of the video is the maximum SI of the frames. Whereas, the TI is calculated between successive frames and hence can be used to represent the change in pixels with respect to time. Both SI and TI evaluates content complexity in terms of space and time respectively. Figure 1 shows the SI and TI of all the videos used in this paper.

![SI and TI of Games Under Test](image)

The SI and TI of GTA V videos is higher than the CSGO and FIFA. This is because in GTA V the adjacent frames are very different and hence the TI was higher than the other games. This was one of the reasons we classified GTA V as a fast-paced game. The content complexity of FIFA was lowest and this results in lower values of SI and TI for most videos tested. This is because most of the frame content are of the grass pitch observed from broadcast camera angle. So, the content of the frames is simpler and results in low SI. Additionally, only a very small proportion of the frame changes with time, resulting in a lowest TI among other GuT. On the other hand, the content complexity of CSGO is between FIFA and GTA V. This is due to the first-person perspective that can result in faster changes within adjacent frames giving it higher TI than FIFA. Moreover, since the user is walking/running and not driving the changes are slower than GTA V and hence this game has a lower TI than GTA V.

IV. QoE EVALUATION USING SUBJECTIVE TESTING

A. Experimental setup

Subjective testing was carried out to determine the QoE of three GuT at different network condition. In this paper, we used 3 different types of network scenarios that are: 1) Delay only, 2) PLR only and 3) Mixed (Delay + PLR). A Linux based emulator NetEm was used to configure different network scenarios shown in Table II.

The cloud gaming scenario was replicated using a three PC setup. The machine designated as cloud (M1) has higher
TABLE I: Summary of network scenarios

<table>
<thead>
<tr>
<th>QoS parameters</th>
<th>Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delay (ms)</td>
<td>10, 20, 30, 50, 75, 100, 200, 300</td>
</tr>
<tr>
<td>Packet Loss Ratio (%)</td>
<td>0.10, 0.25, 0.50, 1, 2, 5, 10</td>
</tr>
<tr>
<td>Mixed (Delay(ms)</td>
<td>Loss(%) )</td>
</tr>
</tbody>
</table>

Table III shows game parameters that were used in this study.

B. Subjective testing and user profile

In this paper, subjective testing was performed to measure the QoE. 22 subjects were employed. Subjects were tested for visual acuity using Snellen chart and none of the subjects reported any visual impairments. The mean age of the subjects was 27.3 years, median age of 28 and standard deviation of age of 4.5. All the subjects had general knowledge of operating computers and were familiar with gaming. There were 4 females and 18 males in this investigation.

Before the start of the experiment the subjects were told about the general mechanics of the game. These include the rules, general gameplay, and control instructions. In addition to that, a short time was provided to get used to the game. The subjects played a game on M2 that had an ASUS 24” (1920x1080p) FHD display. The distance between the user and the screen was maintained according to [10] and the lighting of the room was regulated according to ITU-T recommendation BT.500-13 [13]. Each subject was asked to play each setting of the network scenario per game for at least 60-90 seconds in accordance to ITU-T recommendation P.809 [14]. Each subject tested 38 different network conditions per game that took them around 60-90 minutes to complete. One game was tested in one sitting to avoid subject fatigue.

After completing each scenario, the subject was asked to rate the Video-Quality (V-QoE), Gameplayability-Quality (GP-QoE) and the Overall-Quality (O-QoE) on a 7 point Absolute Category Rating (ACR) scale as recommended in [14]. The V-QoE demonstrates the graphics of the game and the video quality perceived by the user, whereas GP-QoE indicates the game response to the user inputs and ease at which game can be played (this is not the user’s ability to play game). Finally, O-QoE is the overall perception of the game for that specific scenario. This is not the mean of V-QoE and GP-QoE.

V. Results

The results obtained by subjective testing were used to calculate the MOS of each game at different network scenarios. We calculated MOS of V-QoE, GP-QoE and O-QoE at each network scenarios. Moreover, we presented our results in three classes of network scenarios that are: 1) Delay only, 2) PLR only and 3) Mixed Scenario (Delay+PLR).

A. Delay only scenario

As expected, the QoE of all GuT shows a downward trend for increasing delay. The consistency in the trend is seen for all different types of quality, see Figure 2. However, the QoE of all three GuT is affected differently for increasing magnitudes of delay. For instance, FIFA performed significantly better (up to 53%) than the other two GuT for all QoE types, whereas it is evident from Figure 3 that GTA performed worst for V-QoE. Moreover, CSGO and GTA have similar GP-QoE and O-QoE for increasing delay.

This can be explained using the content complexity and the pace of the game. FIFA has the lowest SI & TI and is relatively slow-paced among all GuT, see Figure 1. This means that in comparison to the other games the consecutive frames in FIFA have higher similarity (lower TI). Hence, even at higher delay V-QoE of FIFA is less affected than other GuT. Moreover, GTA has higher content complexity (SI & TI) and thus resulting in a lower V-QoE compared to other GuT for increasing delay.

Additionally, we found that G-QoE depends upon the pace of the game, level of user interactivity and the gameplay of the game. For instance, in FIFA the user spends considerable amount of time in slow paced actions such as passing and dribbling. This means that the user has more time to register moves and thus delay has lower impact on game-playability of slowed paced games such as FIFA, whereas CSGO and GTA have a faster gameplay and require quicker action such as aiming and driving. This means the user needs to interact/react quicker to changing scenarios of the game. In this case, a small delay at a crucial moment of the game can result in reduced GP-QoE.

Another interesting observation was that the O-QoE follows GP-QoE closely for delay scenarios for all GuT. This can be explained with respect to the level of interactivity and...
immersion in the game that is experienced by the user. For instance, in decisive moments of the gameplay subjects showed inclination towards game-playability rather than the video quality. Consequently, when the video quality is bad, but the user is still able to play fine, the user tends to rate the overall quality higher. The opposite of this user scoring behaviour was rarely seen in this investigation.

B. PLR only scenario

Similar to the delay-only scenarios, the QoE values of all GuT show a downward trend for increasing PLR. The consistency in the trend is seen for all different types of QoE, see Figure 3. The QoE of all three GuT is affected very similarly by the increasing levels of PLR. However, GTA still performed the worst out of the three GuT for all types of QoE and at PLR>1.

An interesting finding was about the V-QoE of all games for increasing PLR. The results show more significant difference than the results reported in literature. As it is reported in [4] that for PLR>1% the V-QoE of the FPS shooting game (like CSGO) degraded heavily. In contrast, all games including CSGO showed a lot of resilience to PLR >1% up till 10%. This can be explained using error corrective mechanism employed by the games/platforms. The improved error concealment methods such as frame interpolation can reconstruct part of the missing frames with significant improvement. This means that even when a specific part of the frame is lost, it is reconstructed by using other available parts of the frame.

The GP-QoE indicate a similar trend to V-QoE for all games but the GP-QoE did not reach MOS=1, even at PLR=25%. GTA V game-playability degraded the most and a high proportion of subjects reported occasional loss of control at scenarios with PLR>5%. For FIFA the GP-QoE decreases more in comparison to the V-QoE for increasing PLR. This is due to the jerkiness and blockiness introduced by the PLR to the video quality that makes it difficult for the users to control and coordinate the ball between the players. We found that for PLR>5% the GP-QoE was highly affected due to video artefacts. In [5], it is reported that FIFA’s video quality degraded sharply for increasing PLR levels. Whereas, for CSGO the GP-QoE shows a very similar result to V-QoE.

Another interesting finding was that O-QoE of all the games did not follow specific other quality factor (V-QoE or GP-QoE). This can be seen in Figure 4 that for PLR<10% the O-QoE follows the G-QoE closely as the user perception of the overall quality was more affected by the ease of the play and the interactivity of the game. Whereas for PLR >10% the O-QoE follows the V-QoE more closely. This shows that when the V-QoE was bad, but the user was still able to interact with the game, users tend to rate the O-QoE lower than the subsequent GP-QoE.

Fig. 2: Different QoE types against Delay-only of GuT

Fig. 3: Different QoE types against PLR-only of GuT
C. Mixed (Delay + PLR) scenario

We report in this section the results of the evaluation of the three types of QoE for mixed scenario i.e., combination of delay and loss scenario. Seven levels of PLR conditions were evaluated for each 10ms, 30ms and 50ms delay, as reported in Table I. The results are presented as heatmaps (see Figures 4, 5, 6), where the colour of the block shows MOS at that network parameter combination. A darker colour shows a MOS closer to 1 (extremely bad quality) and a lighter colour shows a MOS closer to 7 (ideal quality).

To elaborate the findings, we also present V-QoE (MOS) as Good to Best (%G2B) and Poor to Worse (%P2W) ratios in Table IV. Where, %G2B is the number of network condition at which the MOS > 4 out of all mixed network scenarios. The same is true for %P2W for MOS<3

TABLE IV: Summary of %G2B and %P2W of different QoE types for GuT

<table>
<thead>
<tr>
<th>QoE type</th>
<th>GuT</th>
<th>%G2B</th>
<th>%P2W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video</td>
<td>GTA</td>
<td>48</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>CSGO</td>
<td>57</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>FIFA</td>
<td>67</td>
<td>19</td>
</tr>
<tr>
<td>Game Playability</td>
<td>GTA</td>
<td>52</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>CSGO</td>
<td>57</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>FIFA</td>
<td>62</td>
<td>24</td>
</tr>
<tr>
<td>Overall</td>
<td>GTA</td>
<td>52</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>CSGO</td>
<td>57</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>FIFA</td>
<td>62</td>
<td>24</td>
</tr>
</tbody>
</table>

1) Video QoE: As expected, for all three games, the V-QoE decreases with increasing delay and loss conditions, see Figure 4. Even in mixed conditions, GTA performed worst among all GuT for every mixed scenario. Similarly, the V-QoE of FIFA degraded least with CSGO in between other two GuT. This clearly indicates how the content complexity plays a vital role in V-QoE of cloud gaming applications. V-QoE of games with higher content complexity are affected most by worsening QoS conditions.

![Fig. 4: Heatmap of V-QoE for mixed network scenarios](image)

It is observed that GTA V has low %G2B for V-QoE as compared to CSGO and FIFA. It also has the highest %P2W among the three games. This confirms that V-QoE of GTA was most affected by the combination of delay and PLR. FIFA has the highest %G2B and lowest %P2W, showing that V-QoE of FIFA was least affected by the network impairments among the other games. In addition to this, CSGO V-QoE is intermediary between GTA V and FIFA.

2) Game QoE: Once again, GTA V has the lowest GP-QoE out of all three games for all network scenarios, see Figure 5. This can be explained due to the pace of the game and hence worsening network condition affects user response more significantly. Unlike V-QoE FIFA showed lower GP-QoE at higher levels of delay and PLR in comparison to CSGO.

This is seen in the second row of Table IV, where %P2W of FIFA is 24% and CSGO shows improvement with a %P2W of 19%. In contrast, at lower levels of delay and PLR FIFA has a better GP-QoE than CSGO. This is observed in Table IV where, FIFA has %G2B of 62% as compared to 57% for CSGO.

3) Overall QoE: Finally, the O-QoE of the GuT was evaluated. All games showed a degradation in O-QoE for increasing levels of QoS parameters, see Figure 6. GTA V showed the lowest at all network scenarios while CSGO and FIFA showed conflicted results at lower and higher magnitudes of PLR as seen in GP-QoE. We used %G2B and %P2W of all three QoE type to evaluate which one out of V-QoE or GP-QoE closely correlate to the O-QoE.

It is observed that for %P2W O-QoE shows similar trend to %P2W seen in GP-QoE for all games. This is also seen at the lower values of delay and PLR (using %G2B). The value of %G2B of GP-QoE is closer O-QoE for all games. This confirms that game playability is a stronger influencing factor than video quality of overall quality perceived by the user. These findings are critical for QoE evaluation of cloud gaming applications and need to be evaluated further.

VI. Conclusion

In this paper, we presented three different types of QoE that can be useful to evaluate the QoE of cloud gaming applications.
applications. We conducted subjective testing to measure the QoE of games at various network scenarios that include delay-only, PLR-only and mixed scenario. We reported our findings and evaluated that GP-QoE relates to the O-QoE more closely than v-QoE for delay only and mixed scenario.

We also found that different network condition affects different types of games differently. We reported that the slow-paced games such as FIFA are less affected by the delay-only but significantly degrades for larger magnitudes of PLR. In addition, we found that games with high content complexity, faster pace and high graphics performed worst at changing network conditions. This was demonstrated using GTA V which consistently performed worst out of all contenders at all network conditions. For medium paced games with medium levels of content complexity such as CSGO a lot of variations was seen in their GP-QoE and O-QoE for different network conditions. These findings are significant for the service providers to tune the network parameters above the threshold level to avoid customer annoyance. These findings are also significant for the game developers and cloud gaming platforms to adapt their infrastructure to tackle changing/degrading network conditions.

In future, we want to extend our analysis further to work out the weights of V-QoE and GP-QoE towards the O-QoE of cloud gaming applications. In addition, further probing is required to determine the key performance indicators (KPIs) of cloud gaming and then correlate them to the subjective QoE, to formulate objective QoE metric for cloud gaming.

REFERENCES


