Online Game Quality Assessment

Luca Venturelli
Amazon
luca.venturelli15@gmail.com

Dorel Picovici
Institute of Technology Carlow
dorel.picovici@ITCarlow.ie

Joseph Kehoe
Institute of Technology Carlow
joseph.kehoe@itcarlow.ie

Abstract

This paper describes an objective model for measuring online games quality of experience (QoE). Most of the existing game quality assessment models take into consideration mainly network impairments, therefore the measured games quality of experience is only correlated with network impairments. In addition to the traditional network-based parameters such as delay, jitter and packet loss, the model described in this paper is using new parameters based on subjective assessment. The aim is to estimate game quality as perceived by an online game player. In order to validate and calibrate the proposed model a subjective game quality assessment is also developed. Two 5-point scales are introduced: a game-quality scale and a game playing-effort scale. The mean average of each scale, termed the Mean Opinion Score (MOS), will indicate the experience of game quality. Reported evaluation results indicate a high level of correlation. This work extends previous work in this area by including end-user knowledge in the proposed metric and demonstrating a high correlation between the obtained measurements and players subjective opinion.

1 Introduction

The focus on distributed interactive applications and the adaptation of the application at both client and server has changed the way multiuser online games are designed and developed. In order to maintain a large number of simultaneous players it is becoming essential for the game service providers to estimate game player perception and game performance. On the other hand, advanced information about game playing conditions would allow game players to select different online games, different networks and ultimately different tariffs. Traditionally the end-user perception for online games has been measured using network game quality assessment only [7]. Although they claim to estimate a user perception, the existing models are based on network components which are limited in fully predicting user satisfaction. Research carried out by US Entertainment Software Association (ESA) revealed that the percentage of online game players is as high as 62% of total game players [3]. These statistics indicate that the average game player has been playing games for 12 years. Adult gamers have been playing for an average of 14 years; males averaging 16 years of game play and females averaging 12 years. The average game player is 30 years old and 15% of most frequent game players pay to play online games.

The main goal of this paper is to provide insight into the quality experience of gamers. A more specific goal of this research is the development of an end-to-end quality measurement method that allows us to quantify the perceived quality of online gaming. This paper considers First Person Shooter (FPS) games, in particular Team Fortress 2. The proposed model, which is designed to capture the users quality of experience (expressed as a Mean Opinion Score (MOS)), is in line with the ITU-T network prediction model (E-Model)[7].

Following this introduction section two reviews the most relevant existing metrics. In section three, a new metric is proposed and its testbed implementation is presented. In section four the tests undertaken are described. Finally in section five the obtained results are analysed.

2 Existing Metrics

The network layer is one of the most important aspects to consider when developing a quality metric for online games. The user perception in terms of network awareness is crucial. In [8] the author reveals that there are two possible approaches for discovering player tolerance to network disruption. The first is to build a controlled lab environment to test small groups of players under selected conditions and the second is to monitor player behaviour on public servers over thousands of games. The *Quake3 G-model* is a model proposed by Ubicom [8] that introduces a new benchmark, *OPScore*, or the Online Playability Score, to describe the effects of network impairments on the playability of online games. The authors focus their attention on latency, the average amount of time necessary to transmit information about a players actions, and jitter, the variance of latency. This technique uses measurements of traffic in a realistic home network environment to forecast the playability of online games. The model defines an impairment factor R given by:

$$R = (W_L \times L + W_J \times J)(1 + E) \quad (1)$$

where W_L, the Latency weighting factor, is set to 1 for the test, E is the Packet loss as a percentage of bytes lost and R is the Impairment factor in ms. W_J is the Jitter Weighting factor calculated with the following formula:

$$W_J = (\Delta K/\Delta J)/(\Delta K/\Delta L) \quad (2)$$

where K is the Average frags per minute, L is the latency and J is jitter as defined in RFC 1889 in ms.

A successive approach proposed by Wattimena [9] uses a similar model named the *Quake IV G-Model* to predict the perceived quality of a First-Person Shooter. The authors conducted a number of subjective experiments to quantify the impact of network parameters on the perceived quality of a FPS game. The model proposed enable us to predict a gamers quality rating based on measured ping and jitter values, and it shows a very high correlation with the subjective data. The final score is given by the following mapping function:

$$MOSmodel = -0.00000587X^3 + 0.00139X^2 - 0.114X + 4.37 \quad (3)$$

where: X - the network impairment is defined as:

$$X = 0.104 \times pingAverage + jitterAverage \quad (4)$$

These experiments demonstrate that ping and jitter have a significant negative effect on the subjective MOS, while packet loss goes unnoticed for values up to 40%. In particular the introduction of jitter in the network has a large negative effect on the perceived quality of the game play. The model has been developed for the game Quake IV and was not tested on other games and platforms. A recent paper [4] presents a new metric to measure the quality of experience (QoE). In this paper we present an analysis of the causal relationship among network delay, system inconsistency and QoE. The QoE is divided into responsiveness, precision and fairness. They create a metric to qualify the view inconsistency. A function is built to map from the objective system inconsistency to the subjective QoE property score. The input of this mapping function considers only the relative inconsistency between views that takes into account any built-in time compensation algorithms, because the same network delay can produce different inconsistencies in different games and the same absolute inconsistency value can lead to different impacts in different scenarios. The approaches described above have at least one of the following weaknesses:

- Although they claim to estimate a user perception, the parameters used are based on network components which are limited in fully predicting the user satisfaction; [8, 9, 4]
- Limited subjective testing for a short period of time with a small user pool; [8, 9]
- While a high level of correlation between the subjective and the proposed models is shown, this could be justified only for one defined game in restricted testing conditions; [8, 9]
- None of the models presented presents a parameter to measure the game client and I/O devices except from the MOS model.[8, 9, 4]

The proposed model overcomes these weaknesses by:

- Extending the network based model to include all parameters. Packet loss must be considered in the bandwidth-intensive application environment;
- Including new parameters not proposed or used previously such as the user-based experience/knowledge factors and distortions introduced by user equipment;

- Moving from a game-specific model to a wide range of existing online games (including console based games);
- Using a number of users for subjective testing in line with accepted test models used in telecommunications based scenarios recommended by ITU-T [7];
- Expecting a level of correlation in excess of 95%.

3 Testbed implementation

The tests have been made using the tool netem [6], a package that comes pre-installed with the Linux OS, which is used as a router to control the traffic. In order to simulate a real game environment all the computers were in the same LAN connected through a router (an Ubuntu PC) to the server that hosted the game Team Fortress 2 in a different LAN.

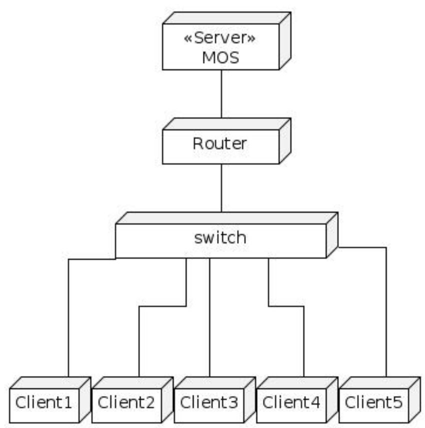

Figure 1: Network infrastructure

Based on the findings in the literature it was decided to simulate the network parameters *delay*, *packet loss* and *jitter*. To emulate the different quality of hardware we performed the test with the following hardware:

1. Dell XPS 8100, 8 GB RAM, Graphic Card Single 1.8GB Nvidia GeForce, Processor Intel Core i7 2.93GHz;
2. Lenovo, 4 GB RAM,Graphic Card ATI Radeon HD 2400XT, Processor Intel Core Duo E4600 2.4GHz
3. Dell Latitude, 8 GB RAM,Graphic Card Nvidia NVS 4200M, Processor Intel(R) Core(TM) i7-2720QM 2.20GHz

The following network tests have been done in order to analyze how the network affects the perceived quality of the game. Table 1 shows the values chosen for the tests on the delay.

Table 1: Network test details

Test Case	**Condition**
1	Delay 0 ms
2	Delay 150 ms
3	Delay 200 ms
4	Delay 300 ms
5	Delay 450 ms
6	Delay 600 ms
7	Packet Loss 15%
8	Packet Loss 30%
9	Packet Loss 45%
10	Packet Loss 60%
11	Jitter 100ms
12	Jitter 200ms
13	Jitter 300ms
14	Jitter 600ms

A series of tests on a range of hardware was performed in order to analyse how the quality of the game changes with different hardware. The tests made were undertaken with the help of 8 players of varying experience (in line with [2]). A second session was subsequently undertaken with different players that combined different hardware with different network quality in order to verify the weight of each parameter. This was implemented to verify the perceived quality with the objective tests. The results are shown in Section 5.

4 Proposed Model

Most of existent game quality assessment models take into consideration only network impairments therefore the measured games quality is only correlated with the network impairments (delay, jitter and, to a limited extent, packet loss) [5]. To estimate the players overall perception of games experience (quality) the model extends the traditional objective game quality methods by introducing the end-user experience/knowledge. As shown in Figure 2 the model uses the following parameters:

1. end-user experience;
2. distortions introduced by game client equipment (memory, graphic card) and I/O devices (screen, keyboard, and joystick);

3. distortions introduced by the network (end-to-end delay, jitter, packet loss);
4. distortions introduced by game server (number of users, game type, game capability to adapt to network distortions).

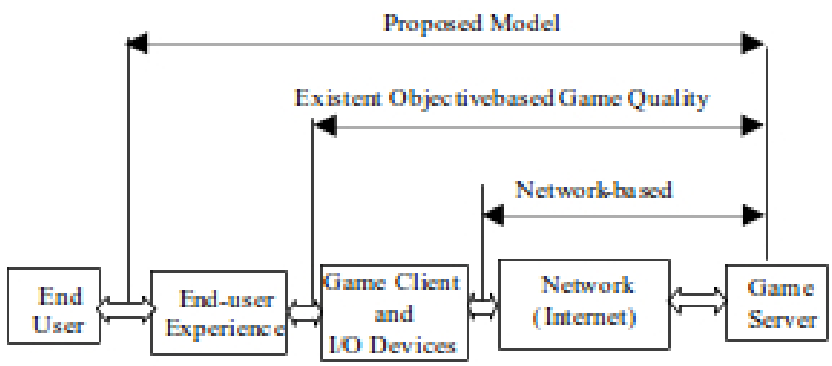

Figure 2: Proposed Model

Using the above-mentioned parameters a Game Rating Factor (GRF) is proposed. The GRF is inspired from an International Telecommunication Union; Telecommunication Standardization Sector (ITU-T) recommended computational model (E-model) [1]. The model is used to assess the combined effects of variation in several parameters that may affect end-user perception of speech quality. The computation of the GRF can be described as follows: a maximum value that reflects the highest level of game quality will be reduced in proportion with the distortions caused by various impairment parameters. Mathematically, GRF can be calculated using the following equation (taken from [1]):

$$GRF = (GRFMAX \times IGCD \times IN + A) \times IGS \quad (5)$$

where each of the terms is defined as follows:

- GRFMAX: maximum Game Rating Factor (90);
- IGCD: impairment factor representing all impairments due to Game Client and I/O device;
- IN: impairment factor representing all impairments due to network connection between the game server and game client;
- IGS: impairment factor representing all impairments due to Game Server (0 or 1);
- A: end-user hands on experience with online games (max 10).

4.1 Impairments due to Game Client and I/O device (IGCD)

To evaluate the Game Client and I/O device, a subjective test was carried out. Data related to game client computer in terms of memory size, processor speed, graphic card, display and mouse were collected and combined as follows:

$$IGCD = 0.4 \times (IGCDmo + IGCDp + IGCDgc + IGCDd) \times IGCDme \quad (6)$$

where:

- IGCDmo: represents the impairment due to the mouse (between 0 and 40);
- IGCDp: represents the impairment due to the processor speed (between 0 and 20);
- IGCDgc: represents the impairment due to the graphic card (between 0 and 30);
- IGCDd: represents the impairment due to the display (between 0 and 10);
- IGCDme: represents the impairment due to the memory of the machine (0 or 1 where 0 is below the suggested memory and 1 is at least equal to the memory suggested for the particular game).

4.2 Impairments due to Network Connection - IN

This factor represents all impairments due to network connection between game server and game client. Data related to game clients network in terms of delay, packet loss and jitter were collected and combined as follows:

$$IN = 0.6 \times (ID + IPL + JI) \quad (7)$$

where :

- ID represents the impairment due to the delay, (value from 0 to 35) ;
- IPL represents the impairment due to the packet loss (value from 0 to 35);
- JI represents the impairment due to the jitter (value from 0 to 30).

The proposed model uses a 5-point scale, see table 2, to measure the quality of an online game. The mean average of the scale, known as Mean Opinion Score (MOS), will indicate the game quality experience (MOS_{GQE}), and mathematically is calculated using the following function:

For $GRF < 0$

$$MOS_{GQE} = 1 \quad (8)$$

For $0 < GRF < 100$

$$MOS_{GQE} = 1 + 0.035 \times GRF + (GRF(GRF - 60) \times (100 - GRF) \times 7 \times 10^{-6}) \quad (9)$$

For $GRF > 100$

$$MOS_{GQE} = 5 \tag{10}$$

The quantity evaluated from the score is represented by the symbol MOS_{GQE} and represents the overall game quality estimation, as perceived by a player.

Wording	**Scale**
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

Table 2: MOS Values

5 Results

A first look at the experiment data shows us that the individual subjective opinions of all the participating players have a high correlation (between 0.74 and 0.95) with the MOS per scenario. This indicates that all of our test users were able to provide more or less consistent opinion scores.

5.1 Factor Analysis

5.1.1 Delay

The influence of delay on MOS is depicted in Figure 3, which graphs the average MOS value against increases in delay duration in milliseconds. The results of our

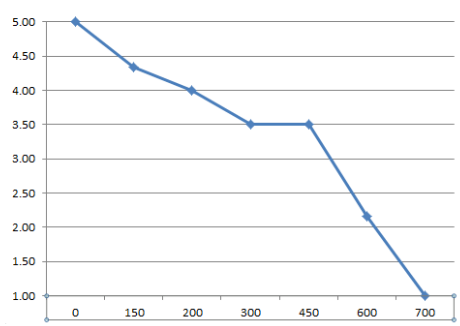

Figure 3: MOS vs. Delay

gaming experiments clearly indicate that higher ping times negatively influence the subjective quality experience of gamers. These results were expected and are totally in line with the outcomes of earlier conducted experiments reported in literature[9].

5.1.2 Packet Loss

The influence of the impairment factor packet loss is not relevant for values under 45% packet loss, as shown in Figure 4. The MOS shows a negative trend when plotted against increasing percentage of packet loss. Packet loss hardly

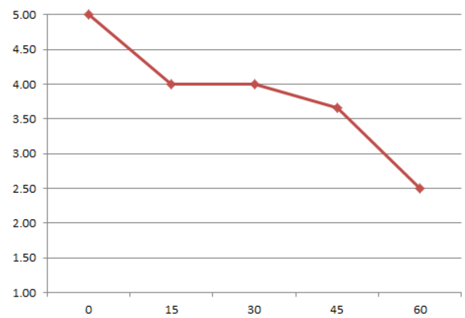

Figure 4: MOS vs. % Packet loss

affects the perceived quality of Team Fortress 2. During the tests we noticed that the number of rendering bugs in the game increases proportianally with the increase of the packet loss.

5.1.3 Jitter

From the tests made we saw that Jitter highly influences the perceived quality of the game. The graph of jitter against average MOS value shows that as the variation in packet latency increases we see that the average MOS value decreases markedly.

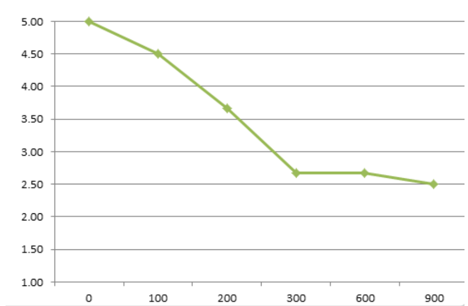

Figure 5: MOS vs. Jitter

5.1.4 Hardware Tests

The results of hardware tests show that the mouse has a very important effect on the perceived quality of the game, especially for more expert players. The quality of the graphic card is the second most important parameter that can affect the game play. The processor doesn't affect the quality of the game too much. Finally we noticed that the quality of the game changes proportionally to the resolution of the screen. Figure 6 shows how the mouse quality affected two different groups of hardware. For the tests we used two Lenovo low quality mouse, two Dell medium quality mouse and for the high quality we choose a Corsair Vengeance M60 and a Roccat Lua Tri-Button Gaming Mouse. The x-axis shows machines with a low quality mouse (1), those with an average quality mouse (2) and finally (3) machines with a high quality mouse.

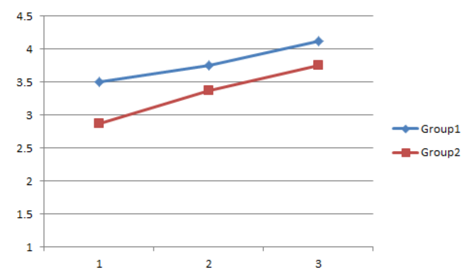

Figure 6: MOS vs. Mouse Quality (3 is better)

As shown in figure 6 the mouse quality has a large effect on the quality of the game. Both hardware groups show similar trends as the quality of mouse used changes.

The next graph (figure 7) shows two different groupings of machines that employ monitors with two different resolutions (marked 1 and 2 on the x-axis).

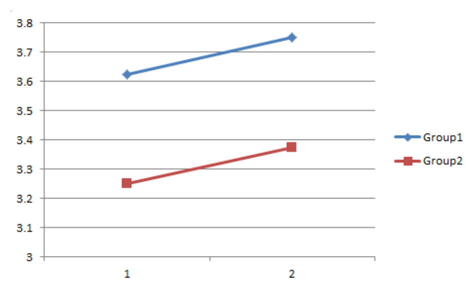

Figure 7: MOS vs. Monitor Quality (2 is better)

As shown in Figure 7 the quality of the monitor impairs the quality of the game but less than the quality of the mouse. We can see that the graphic card has a strong impact on the quality of the game. All measurements of the Netem network degradation scenarios and the hardware degradation scenarios were used as a training set to develop a gaming model for the prediction of the number of the MOS. The correlation R between the subjective tests and the model was 0.97.

6 Conclusion

In this paper we have examined the (simultaneous) impact of ping, jitter, and packet loss on the game-experience for the FPS Team Fortress 2. We also have introduced a new concept, how hardware can affect the quality of the game. Our experiment results demonstrate that delay and jitter have a significant negative effect on MOS, while packet loss goes unnoticed for values up to 45%. The introduction of jitter in the network had a large negative effect on the perceived quality of the Team Fortress 2 players. Mouse and Graphic card are the most important hardware components that affect the quality of the game.

Making use of a multi-dimensional regression analysis we have proposed a new metric which enables us to predict a gamer's quality rating based on different parameters. The correlation between the subjective data and our model is more than 95%. A follow-up validation experiment with different hardware and different network qualities showed that the model is very accurate in estimating these MOS values. In the near future we plan to validate how well this model performs for other online games such as strategy games.

References

[1] ITU-T Recommendation G. 107. The e-model, a computational model for use in transmission planning. March 2003.

[2] Grenville Armitage and Lawrence Stewart. Some thoughts on emulating jitter for user experience trials. In *Proceedings of 3rd ACM SIGCOMM Workshop on Network and System Support for Games*, NetGames '04, pages 157–160, New York, NY, USA, 2004. ACM.

[3] Entertainment Software Association. 2012 essential facts about the computer and video game industry. *http://www.theesa.com/*, 2012.

[4] P. Chen and M. El Zarki. Perceptual view inconsistency: an objective evaluation framework for online game quality of experience (qoe). *Proceedings of NetGames 2011, ACM SIGCOMM, Ottawa, Canada*, 2011.

[5] Tanja Lang, Philip Branch, and Grenville Armitage. A synthetic traffic model for quake3. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '04, pages 233–238, New York, NY, USA, 2004. ACM.

[6] netem, as of April 2013. http://www.linuxfoundation.org/collaborate/ workgroups/networking/netem.

[7] ITU T Recommendation P.800. Methods for subjective determination of transmission quality. *http://www.itu.int/rec/T-REC-P.800-199608-I/en*, 1996.

[8] Ubicom White Paper. Opscore or online playability score: A metric for playability of online games with network impairments. *http://www.ubicom.com/news/whitepapers/ whitepapers.html*, 2005.

[9] A.F. Wattimena, R.E. Kooij, J.M. Van Vugt, and O.K. Ahmed. Predicting the perceived quality of a first person shooter: the quake iv g-model. *Proceedings of NetGames 2006, ACM SIGCOMM, Singapore*, October 2006.