Application of QoE evaluation methodology

A study of the user's acceptance threshold regarding image quality

Graziella Cardoso Bonadia
Maria Silvina Medrano
Nyvea Maria da Silva
Service Technology Corporate Department
CPqD Foundation
Campinas, Brazil
{bonadia, mmedrano, nsilva}@cpqd.com.br

***Abstract* — Taking into consideration the methodology to evaluate the user's quality of experience (QoE), the objective of this work is to evaluate the video image acceptance threshold regarding its compression ratio. The aim is to connect technical and subjective assessments gathered directly from the user. During this first phase, aspects regarding the transport network failure were evaluated. The ultimate goal of this work in progress is to obtain the technical parameters associated with the QoE, so that the best setup is used when adding videos to the Digital TV interactive services.**

Keywords-Quality of Experience; compression ratio; video; network condition.

I. INTRODUCTION

The objective of video compression systems is to reduce the bit rate required for transmission or storage with enough quality for the application. In general, two techniques are used: the exploration of the psychovisual properties of the human visual system and the reduction of the statistical redundancy in images (spatial and temporal). Depending on the efficiency of the compression algorithm used, some kinds of degradations are introduced in the video signal, which can be perceptible or not to the end user. In general, these degradations do not occur separately and the final quality perceived by the user will be affected through the combination of all the effects since the creation of the content, its encoding, packaging, transport up to the reception and display on the user's device. Due to all the technical variables influencing the image quality, the user acceptation is not necessarily determined by each of these factors. Although the technical evaluation has great importance on the final quality, it is the user’s attitude towards the service (video on demand, streaming, among others) that will determine the proper level of quality. There are many studies regarding technical aspects of image quality ([1] and [2]). Those studies carry interesting and successful methodologies to assess technical variables on image. But on a market point of view, the best quality is not always requested. The best quality is not always perceived by end users and their previous experiences about a particular service. To include user’s perceptions on assessment process provides information on balance between the “best” and the “enough”, optimizing the use of resources to deliver a service. As stated in [3], ‘methods to assess the visual quality of digital videos as perceived by human observers are becoming increasingly important…’. This study takes into consideration the user opinion about quality of images. The concept of Quality of Experience has taking much attention for all kind of service and it is being used for a variety of issues ([4], [5] and [6]).

Quality of Experience (QoE) can be defined as the acceptance of an application or service subjectively perceived by the user regarding performance and usefulness, including system components (terminal, network, service infrastructure, etc.), as well as context of use and end-user expectations [7]. Nevertheless, the assessment of services that takes into account the user’s opinion is largely based on a single score ([3] and [8]). When assessing a service, it must take into consideration that people have different expectations and they translate these expectations in distinct manner for each distinct service. For this reason, this paper will consider a methodology that combines two answers from users: The score on a well known five point scale and the binary answer about whether the service (or video image in the case) is acceptable for use on a routine basis. This approach will provide a reference curve of score and acceptability which accounts for the target market and the assessed service.

Considering the methodology to evaluate the user's QoE [9] based on laboratory tests, the purpose is to evaluate, throughout several video sessions, the threshold of the technical variables values that will be perceived by the user, and which are not accepted by the market.

The typical procedure for subjective evaluation consists of submitting to a group of evaluators a sequence of video streaming within a controlled environment. Based on these sequences, the evaluator is asked to give an opinion on the quality and acceptance in watching a program with the same image quality level. After the opinions are gathered, a statistical analysis is performed resulting in numbers for comparison.

The purpose of the work in progress is to evaluate, at first, different video compression ratios (considering the same codec used for all five cases) in the viewpoint of the user's QoE. The following steps will be to introduce simulations regarding the network quality in the user's perception. This paper analyzes two network quality cases (simulated on laboratory) in addition to five compression ratios. The ultimate goal for future work is to obtain data based on different types of images with five video compression ratios and through several packet loss

 ISBN: 978-1-61208-232-5

simulations (in order to simulate network conditions), so that a threshold value is obtained regarding the users acceptance.

At first, the compilation of scores and the acceptance regarding the image quality are performed in order to evaluate the user's QoE. Then, these values are submitted to a logistic regression, with categorical variables (see, for instance, [10]) regarding acceptance probability evaluation according to the score. Based on this set of probabilities, the linear regression between the scores and the probabilities is performed in order to obtain the reference curve. In this methodology, the final value of the QoE is the probability obtained from the data average score of 50% of the most positive scores. As a standard, only probabilities above 0.7 will be deemed acceptable. In the following sections, the laboratory test conditions, the analysis and the preliminary results will be presented.

II. TEST PREPARATION

During the first phase, where the objective was to evaluate separately the user's perception regarding the image quality according to the video compression ratio, simulation preparation procedures were performed to be presented to the user.

First, nine 11-second images were selected with different levels of complexity, such as movement, landscape, drawings, etc. Each image went through five video compression ratios: 3 Mbps, 6 Mbps, 9 Mbps, 12 Mbps and 15 Mbps. Furthermore, each compression ratio condition also passed through a packet loss to simulate transport network failure. As a result, 90 test conditions would be evaluated. Every single test condition was evaluated at least by ten users. The video room was fitted with a 42" TV and three seats for the evaluators at a distance of three meters from the TV set. As a whole, ten 15-minute test sessions were performed with three evaluators each, as well as a number of random image conditions (see design of experiments methodology by [11]). Each evaluator was given a questionnaire with the following questions:

- What score would you give image 1: 5, Excellent; 4, Good; 3, Average; 2 Poor; or 1, Very poor?
- Would you watch a program with this image quality?

The evaluators were requested to answer the two questions for every displayed image.

The data for analysis were obtained only after the ten laboratory test sessions were performed, with each condition repeated ten times. The purpose of the different complexity levels of the images was to obtain the average score of the displayed videos that simulate standard TV programs. For this reason, they were not considered as variables to be studied. The variables considered for user QoE analysis include the compression ratios divided in five levels and the packet loss divided in two: with and without loss. The following table illustrates the analysis.

TABLE I. CONFIGURATION OF PERFORMED TESTS

Compression Ratio	***Without packet loss***	***With packet loss***	***With and without loss***
3 Mbps	Yes	Yes	Yes
6 Mbps	No	No	Yes
9 Mbps	No	No	Yes
12 Mbps	No	No	Yes
15 Mbps	Yes	Yes	Yes

(*) Although all types of combinations were calculated, only the extremes are analyzed in this paper.

The analyses presented in the next section are:

A comparison between the QoE estimate at a compression ratio of 3 Mbps and 15 Mbps (both with and without packet loss): the main goal is to evaluate whether packet loss simulation was perceptible to the user so that other simulations are created with more loss conditions during the next tests (higher or lower than the current test).

A comparison between the compression ratios, considering both packet loss conditions: the objective is to evaluate the compression ratio separately and to determine the ratio perceptible to the user in terms of QoE.

III. ANALYSIS

The QoE analysis is connected to the users' perception on the use of technology. Thus, the purpose of the QoE estimate methodology used is to provide a value that would quantify the subjectivity of the user evaluation as a form of acceptance probability threshold. As stated previously, the evaluator was requested to give a score according to the pre-established scale and then determine if the image quality is acceptable for a standard TV program. The combination of both will be used to adjust the scale and provide an answer suitable for the target audience.

Through an analysis based on model adjustments, the scores and the acceptance status of the service are submitted to a logistic regression, followed by a linear regression that results in a reference curve (given its characteristics). This value, which varies between 0 and 1 necessarily, is the QoE estimate. This value structures market aspects and provides clear data on the number of people within the target market that would watch the program within the evaluated format.

Equation 1 shows the logistic regression adjustment:

$$ln\,(p/(1-p)) = a + b*score(1) + c*score(2) + d*score(3) + e*score(4) + f*score(5) \qquad (1)$$

where:

In p/(1-p) is the acceptance or non-acceptance of the image quality;

p is the acceptance probability, for a given *score(x)*;

a, b, c, d, e, f: are the parameters adjusted by the logistic regression;

score(x): is equal to 1 if the score is *x* or *0*, otherwise.

Equation 2 shows the acceptance probability:

$$p= e^{\hat{a}+\hat{b}x}/(1+e^{\hat{a}+\hat{b}x}) \qquad (2)$$

where *b* is the parameter for *score(x)*.

After calculating the probability for each score and each test condition, the reference curve was adjusted by the linear regression. The results are presented in two parts. Firstly, the analysis is performed considering the comparison between the images with the five compression ratio values, without distinguishing whether or not there is a packet loss. Then, another analysis is performed to assess if the procedure used to simulate the packet loss is perceived by the user. To do so, the video compression ratios at 3 Mbps and 15 Mbps with and without packet loss are compared. Table II lists the probability values according to the reference curves in both analyses:

TABLE II. ACCEPTANCE PROBABILITY BY SCORE – ANALYSIS 1

	Analysis 1				
Score	*3 Mbps*	*6 Mbps*	*9 Mbps*	*12 Mbps*	*15 Mbps*
1	0.04	0*	0*	0*	0*
2	0.24	0.20	0.2	0.17	0.19
3	0.45	0.46	0.45	0.44	0.44
4	0.65	0.71	0.69	0.70	0.69
5	0.86	0.96	0.95	0.97	0.93

(*) Minimum and maximum approximate values.

TABLE III. ACCEPTANCE PROBABILITY BY SCORE – ANALYSIS 2

	Analysis 2			
Score	*3 Mbps without packet loss*	*3 Mbps with packet loss*	*15 Mbps without packet loss*	*15 Mbps with packet loss*
1	0.12	0*	0*	0*
2	0.29	0.22	0.17	0.21
3	0.46	0.45	0.41	0.48
4	0.63	0.68	0.65	0.74
5	0.80	0.9	0.89	1*

(*) Minimum and maximum approximate values.

After the calculation, the reference curves were adjusted to the five compression ratio cases and the four combinations between the compression ratios and the packet loss status. The purpose of reference curves is to associate the given score with the QoE estimate in cases when scores given by the users are still required, even after the product is launched. Figure 1 illustrates the five adjusted reference curves, while Figure 2 shows the four curves related to the packet loss status.

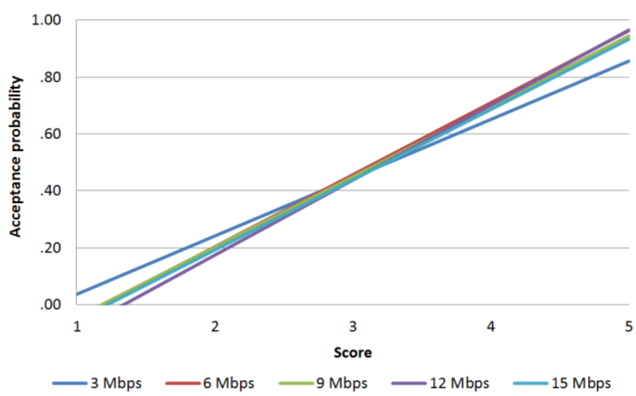

Figure 1. Reference curve – Analysis 1

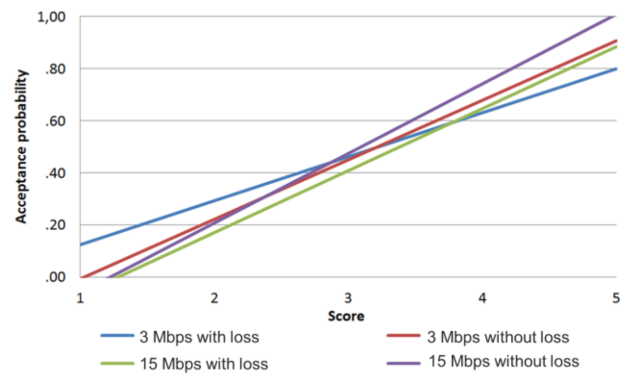

Figure 2. Reference curve – Analysis 2

In this context, the QoE estimated is the probability obtained from the curve regarding the average score of 50% of the data compiled during the laboratory tests. Figure 3 shows the QoE estimate of each compression ratio setup.

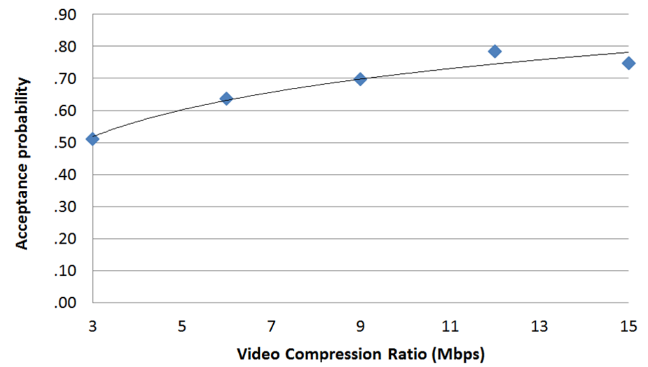

Figure 3. Quality of Experience estimate

As shown in Figure 3, the QoE level increases according to the compression ratio, which means that the user is aware of the difference in quality between the images. However, the perception is decreased when considering compression ratios at 12 Mbps and 15 Mbps. The QoE values are established by the probability of a person to agree to watch a program with the image quality under evaluation. Furthermore, the probability for the person to accept the service, considering that he/she belongs to the 50% potential users, is 0.75 for the 15 Mbps compression ratio and 0.78 for 12 Mbps. Within the acceptance criteria, the user considers the images with 9 Mbps compression ratio to be of good quality, even if the difference with better ratios is perceptible.

To evaluate if the criteria to run the network interference simulation, with random packet loss throughout the video, cause changes in the image quality, tests were performed comparing high-quality and low-quality images, with and without packet loss. Figure 4 shows the QoE estimate for each case. Simulation 1 represents cases with no packet loss, whereas Simulation 2 represents packet loss.

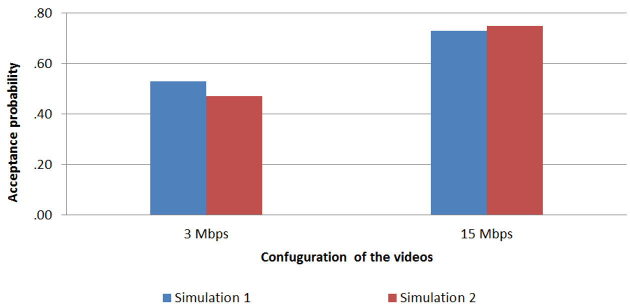

Figure 4. Quality of Experience estimate

The analysis shows that the difference obtained from the comparison is due to the compression ratio. When this characteristic is discarded, 3 Mbps image quality is perceived less in Simulation 2 than in Simulation 1. However, when performing the chi-square test to evaluate whether there was a significant difference, the result was 0.99. By comparing the calculated value (0.99) with the value from the chi-square table, considering 5% significance (3.84), the result indicates that there is no evidence the packet loss simulation is perceived by the user regarding image quality. On the other hand, by comparing the compression ratios, both with packet loss, the calculated value of the chi-square test was 12.11, much higher than 3.84. This means that the image quality level is perceived by the user.

IV. Final Considerations

At this fist part of the project, the analyzed data was used to evaluate the user perception mainly regarding the different video compression ratios, considering a group of images with different levels of complexity. The objective was to evaluate if the user perceived the difference in the image quality and when it would be acceptable. The results obtained from the analysis showed that there is a perception and the images with compression ratios above 9 Mbps are accepted, according to the established criteria. Nevertheless, the difference between the 6 Mbps and 9 Mbps compression ratios is considerable only with 10% significance level. Images with 3 Mbps compression ratio are substantially distinct from all other conditions.

To evaluate the packet loss status, a specific methodology for randomized image degradation was used. The result was not perceived by the user. Therefore, to evaluate the tolerance thresholds on the network aspects, new simulations with specific packet loss techniques with a larger scope than the one used in this study must be performed.

Similarly, other variation sources will be the object of analysis in future studies. Examples of variation sources are: network parameters and the different sizes and technologies used in the receivers. The objective is to obtain a response surface in order to evaluate how the set of parameters contributes to a tolerance threshold or acceptance of a program in the user's perception, translated in QoE estimates.

References

[1] Z. Wang, L. Lu, and A. C. Bovik, "Video Quality Assessment Based on Structural Distortion Measurement", Signal Processing: Image Communication, vol. 19, no. 2, Feb. 2004, pp. 121-132.

[2] C. Yim and A. C. Bovik, "Evaluation of Temporal Variation of Video Quality in Packet Loss Networks", Signal Processing: Image Communication, vol. 26, no. 1, Jan. 2011, pp. 24-38.

[3] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, " Study of Subjective and Objective Quality Assessment of Video", IEEE Transactions on Image Processing, vol. 19, Issue: 6, June 2010, pp. 1427-1441.

[4] I. Wechsung, K. P. Engelbrecht, C. Kühnel, S. Möller, and B. Weiss, "Measuring the Quality of Service and Quality of Experience of Multimodal Human-Machine Interaction", Journal on Multimodal User Interfaces, vol. 6, Issue 1-2, July 2012, pp. 73-85.

[5] H. A. Tran and A. Mellouk, "QoE Model Driven for Network Services", Wired/Wireless Internet Communications – Lecture Notes in Computer Science, vol. 6074, 2010, pp. 264-277.

[6] D. K. Moore *et al*., "Linking Users' Subjective QoE Evaluation to Signal Strength in an IEEE 802.11b/g Wireless LAN Environment", EURASIP Journal on Wireless Communications And Networking, vol. 2010, pp. 1-12, Article ID 541568, doi: 10.1155/2012/541568.

[7] ITU, Recommendation G.1080. "Quality of experience requirements for IPTV services", 2008.

[8] J. Xu, L. Xing, A. Perkis, and Y. Jiang, "On the Properties of Mean Opinion Scores for Quality of Experience Management", IEEE International Symposium on Multimedia (ISM – 2011), Dec. 2011, pp. 500-505, doi: 10.1109/ISM.2011.88.

[9] G. C. Bonadia, C. A. Tambascia, and E. M. Menezes, "Standardizing the Evaluation of QoE by Users – A Methodology to Estimate the Acceptance of Interactive Services", The Fifth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies and Services (CENTRIC 2012), Nov. 2012, in press.

[10] D. W. Hosmer and S. Lemeshow, Applied Logistic Regression, 2nd ed., New York: John Wiley & Sons INC., 2000.

[11] D. C. Montgomery, Design and analysis of experiments, 8th ed., New York: John Wiley & Sons INC., 2012.