Quality Metric to Assess Video Streaming Service over TCP considering Temporal Location of Pauses

Demóstenes Z. Rodríguez, Julia Abrahão, Dante C. Begazo, Renata L. Rosa, and Graça Bressan

Abstract — *There is a wide range of video services over complex transmission networks, and in some cases end users fail to receive an acceptable quality level. In this paper, the different factors that degrade users' quality of experience (QoE) in video streaming service that use TCP as transmission protocol are studied. In this specific service, impairment factors are: number of pauses, their duration and temporal location. In order to measure the effect that each temporal segment has in the overall video quality, subjective tests. Because current subjective test methodologies are not adequate to assess video streaming over TCP, some recommendations are provided here. At the application layer, a customized player is used to evaluate the behavior of player buffer, and consequently, the end user QoE. Video subjective test results demonstrate that there is a close correlation between application parameters and subjective scores. Based on this fact, a new metrics named VsQM is defined, which considers the importance of temporal location of pauses to assess the user QoE of video streaming service. A useful application scenario is also presented, in which the metrics proposed herein is used to improve video services[1].*

Index Terms — Video Streaming, TCP, Video Quality, Subjective Test, QoE, MOS.

I. Introduction

Many video services over the Internet have recently gained popularity, and for this reason video traffic is increasing considerably. Due to this huge demand for video services, the study of methods that assess video quality and therefore user quality experience (QoE) is very important. Internet video streaming service providers must be prepared to serve users in large scale, ensuring an acceptable quality of service [1].

Nowadays, most of these video services run over HyperText Transfer Protocol (HTTP), which uses Transmission Control Protocol (TCP) as the transport protocol. This is because communication services, based on User Datagram Protocol (UDP), are in some cases intercepted and blocked by firewalls or Network Address Translation (NATs), and for this reason, streaming video services based on UDP are not offered frequently.

One problem that arises in a network such as the Internet is congestion. To minimize it, congestion control mechanisms for TCP are implemented [2-4]. When TCP detects packet losses, the TCP transfer rate decreases, and if this new rate is smaller than the playback rate, the player takes all the buffer information and then enters a rebuffering process. During this rebuffering period, no information is displayed and this causes degradation of end user QoE.

In the application layer, this work uses a customized player in order to extract information regarding the player buffer states during video streaming. These states are the application layer parameters and they indicate mainly: number of pauses and their frequency, mean pause length and temporal location.

Results of subjective tests play an important role, because product improvements are based on users' requirements [5]. In subjective testing of video quality, the human perception on the quality of the tested material is quantified by a score and the quality of the service is evaluated according these results.

Here, the most important current subjective test methods to assess video quality are described, highlighting that none of them is appropriate to assess video streaming service running over TCP. As a consequence, some criteria based on the mechanism of human cognitive process are established, considering that the effect of pauses has a different nature from impairments originated, for example, by video compression algorithms. During pauses, no information is presented to users, whereas during other impairments, the information has quality degradation; nevertheless, video sequences are always presented.

Furthermore, it is important to note that objective metrics such as: Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [6], Video Quality Metric (VQM) [7] and algorithms based on Region of Interest (RoI) [8] or attentions maps [9, 10] are not suitable for video streaming over TCP, because they do not take into account the effects and characteristics of pauses.

The most important contribution of this work is to determine a new metrics that quantifies the user QoE in video streaming service over TCP. Subjective test scores are thus correlated with the application layer parameters and, as a result, a new metrics named Video streaming Quality Metrics (VsQM) was defined, and mapped in MOS scale. The new approach took into account the temporal location of each pause. Thus, QoE did not only depend on the number of pauses and their mean period of time, as stated in [11,12]. For video streaming service over TCP, the temporal location of each pause also played an important role.

[1] This work was supported by FAPESP (The São Paulo Research Foundation –Brazil). FAPESP project number: 2011/12724-8.

The authors are with the Computer Engineering Department at the School of Engineering - University of São Paulo, Brazil. (e-mails: {demostenes, gbresan}@larc.usp.br, {jabrahao,rrosa}@usp.br,mailto: and dante@lps.usp.br)

Contributed Paper
Manuscript received 07/14/12
Current version published 09/25/12
Electronic version published 09/25/12.

Once the proposed metrics was defined, a useful scenario was presented where this metric was used. In this scenario, a feedback mechanism was used to send the quality metrics from end user to service provider, which can be used for different purposes such as monitoring, reports or as input of a Rate Determination Algorithm (RDA).

In this context, the remainder of this paper is structured as follows. Section 2 presents the Subjective Test Methods to assess video quality. Section 3 introduces the Proposed Video Quality Model used to determine the VsQM. Section 4 shows the Test Scenario and describes the customized player, the video data base characteristics, and a service application scenario. Section 5 presents the results, indicating the values of each segment weight, and a function which maps the VsQM in the 5-point MOS scale is defined. Finally, section 6 presents the conclusions.

II. Video Quality Subjective Test Methods

ITU-T defines QoE as: the overall acceptability of an application or service, as perceived subjectively by the end user [13]. In [14], this concept is enhanced because it includes objective human cognitive aspects and incorporates some valid psychological subjective and social approaches for the assessment of QoE. Considering all the factors that needed to be included in the latter context, the study needed to be multi-disciplinary, and incorporate disciplines such as: psychology, cognitive science, sociology and information technology.

We studied how the concepts of the mechanisms that integrate perceptual and cognitive processes must be related to a test environment for the subjective assessment of video, which was limited to the service of video streaming transmitted over HTTP / TCP protocols.

Perceptual and cognitive processes occur on a continuum. Each environment stimulus presents a set of characteristics such as color, texture, size, shape, brightness, motion, which are captured by our senses, and this process is called low sensory process. This information is converted into electrochemical stimuli, transmitted to our central nervous system and interpreted, organized, stored and recalled. These treatments occur by different cognitive processes such as memory, categorization, attention, problem solving and decision-making processes, which provide indicators for action. For this reason, it is not adequate to consider only the low-level sensory process to determine the perceptual video quality [15,16].

In this section, firstly, the most popular current subjective test methodologies for video quality assessment are reviewed, and the restrictions of these methods to assess video streaming service transmitted over TCP are highlighted. Secondly, temporal pooling methods for quality video assessment are treated. And finally, some recommendations used to conduct the subjective tests in this work are presented.

A. Current Subjective Test Methodologies for assessing Video Quality

There are several subjective tests methods, most of them described in the ITU recommendations: ITU-R BT-500 [17] and ITU-T P.910 [18], the latter focused on multimedia services. ITU-R BT-500 described the following methods: Double Stimulus Impairment Scale (DSIS), Double Stimulus Continuous Quality Scale (DSCQS), Single Stimulus Continuous Quality Evaluation (SSCQE) and Simultaneous Double Stimulus for Continuous Evaluation (SDSCE). Whereas ITU-T P.910 describes, Absolute Category Rating (ACR), Absolute Category Rating with Hidden Reference (ACR-H), Degradation Category Rating (DCR) and Pair Comparison method **(**PC). The Assessment Methodology for Video Quality (SAMVIQ) is described in [19], and [20] considers some variants of the previous methodologies. Table I shortly presents the main features of the video quality assessment methods most cited in current literature.

TABLE I
Parameters of Video Quality Assessment Methods

Methods	Video Length (s)	Explicit Refe-rence	Hidden Refe-Rence	Simulta-neous Stimuli	Continuous Quality Scale
DSIS	10	Yes	No	No	No
DSCQS	10	No	Yes	No	No
SSCQE	300	No	No	No	Yes
SDSCE	10	No	No	Yes	Yes
SAMVIQ	10	Yes	Yes	No	No
ACR	10	No	No	No	No
ACR-HR	10	No	Yes	No	No
DCR	10	Yes	No	No	No
PC	10	No	No	No	No

As shown in Table I, all these methods, except SSCQE, are oriented to assess the effects of spatial degradation in the video frames, because they consider video length around 10 seconds, being more useful in scenarios where the objective is to evaluate the performance of a particular encoder. Using these methods, the effects of degradation caused by pauses cannot be properly evaluated.

The SSCQE methodology considers a longer duration of video sequence and a continuous assessment by the user. This methodology application, however, would not be reasonable to grant a score during the pause.

B. Temporal Pooling Methods for Video Quality Assessment

In order to pool the distortions in spatial and temporal position, video quality metrics more commonly use different methods; suggest direct averaging [21] or the Mikowoski method. Other studies, such as [22], propose a temporal pooling function, which gives more weight to the end part of video, but how this function is obtained is not explained. Moreover, the ITU-R recommendation BT.500-11 states that "*human memory effects can distort quality ratings if noticeable impairments occur in approximately the last 10-15 seconds of the sequence*", and also indicates that this behavior could be modeled as a decaying exponential function.

Few papers study how temporal degradations affect the score of the final quality of the video. The relation between time-varying quality scores and the final quality score assigned by assessors to a video using a subjective method based on ACR scale and continuous assessment is studied in [21], and the conclusion is that scores assigned by assessors to a video can be approximated to the mean of their continuous quality scores. It is worth noting that, in [21], different distortions such as MPEG-2 and H.264 compression were used, and they simulated transmission errors over IP network. Also, the video length used was 10 seconds, different from the test scenarios considered in our work.

In order to determine whether these temporal pooling methodologies are valid to video streaming over TCP, a set of video tests were evaluated. These videos have the same number of pauses, but these are located in different instants of video sequences.

C. Proposed Subjective Test Methodology

In subjective tests conducted in this work, the variability of the cognitive processes of individuals is considered. This way, assessors have different characteristics, such as: attention degree, speed in processing information, short-term memory and long-term memory in addition to prior knowledge and even preferences about video content. Therefore, the criteria reported in Table II were considered to conduct the subjective tests.

TABLE II
SET OF CRITERIA CONSIDERED FOR CONDUCTING SUBJECTIVE TESTS

Item	Considerations
1	Video sequences were longer than 10 seconds, in order to prevent loss of the effect of pauses temporality. Video sequences with two and four minutes were chosen taking as a reference the top videos in the most popular *video-sharing* service, which have a video length of three minutes.
2	Assessment tests needed to be conducted in an environment similar to the actual use of video streaming services. In the tests, users made the streaming videos following their favorite sequence, and they could watch each video sequence as many times as they deemed necessary.
3	Considering the variability of attention of assessors, the instructions for performing the tests had to be specific. Consequently, at the beginning of the test, the instructor advised assessors that video degradation is only due to the presence of pauses, so that the assessors' attention was focused on this degradation factor.
4	The tests were not conducted in groups of assessors, all test were performed individually, because the evaluations should be performed when the users found that they understood the test procedure. The assessors had different speeds in processing information, hence no limited time to receive their scores was considered.
5	Considering the variability in assessors' memory, they could watch the videos as many times as each of them considered necessary; changing the scores was allowed, as indicated in the SAMVIQ methodology.

In order to rate each video, assessors used a 5-point MOS scale described in ITU-T Recommendation P.910, shown in Table III.

TABLE III
ITU-T 5-POINT SCALE - ACR

Grading Value	Estimated Quality	Perceived Impairment
1	Excellent	Imperceptible
2	Good	Perceptible but not annoying
3	Fair	Slightly annoying
4	Poor	Annoying
5	Bad	Very annoying

III. THE PROPOSED VIDEO QUALITY MODEL

Initially, subjective video tests were conducted following the consideration of Table II. These tests were preliminary, and the conclusion was that a metric for assessing quality video in streaming video service have to consider the temporal location of each pause in the video. Based on this criterion, it was necessary to build several network scenarios to create pauses at different instants of the video and with a certain duration.

Hence, the concept of a video temporal segmentation is introduced herein, and the following segments are defined: (a) segment A, initial video segment; (b) segment B, first intermediate segment; (c) segment C, second intermediate segment; and segment D, final video segment.

The VsQM metric proposed herein was determined by the parameters: number of pauses, pauses length and weight of the temporal segment during which the pauses occur. Fig. 1 helps interpreting the meaning of the proposed metric, at which the video playback time was of T_D seconds; in this scenario, four time segments were considered, with six pauses of different durations distributed randomly. The number of segments could be increased, but to calculate the weight of all segments, more test video sequences would be necessary.

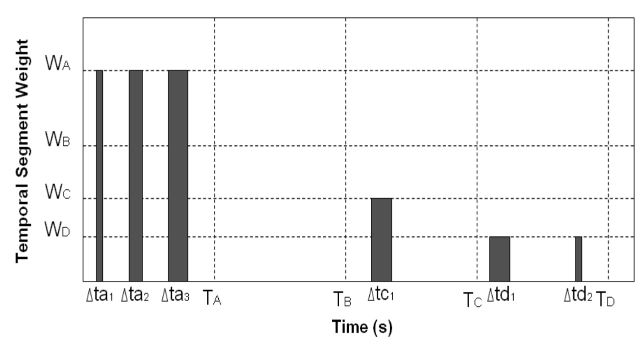

Fig. 1. Parameters that compose the VsQM metric

The VsQM metric is defined by (1):

$$VsQM = \sum_{i=1}^{k} \frac{N_i * L_i * W_i}{T_i} \quad (1)$$

Where:

- N_i is the number of pauses;
- L_i is the average length of pauses, in seconds, that happened in the same temporal segment;

- W_i is a weigh factor which represents the degree of degradation that each segment adds to the total video degradation;
- T_i is time period in seconds of each segment;
- k is the number of temporal segments of a video, this work considered four segments for all the tests.

With the results of subjective video tests, it was possible to determine a weight related to the degree of degradation that each segment (W_A, W_B, Wc and W_D) caused in the total degradation of the video.

Also, in order to map the VsQM values with MOS index values of 5-point scale, an exponential function was taken as a model. VsQM at MOS scale is denoted as $VsQM_{MOS}$, and it is presented in (2).

$$VsQM_{MOS} = C * \exp(-\sum_{i=1}^{k} \frac{N_i * L_i * W_i}{T_i}) \quad (2)$$

In this work were performed twenty different test scenarios, which will be explained in detail in the next section. An MOS index for each assessed scenario resulted from the subjective tests, and these MOS values will be used to model the VsQM, for instance, the following relation corresponds to scenario 1 ($VsQM_{MOS-1}$):

$$Ln(VsQM_{MOS-1}) = Ln(C) - \frac{W_A * N_A * L_A}{T_A} - \frac{W_B * N_B * L_B}{T_B} - \frac{W_C * N_C * L_C}{T_C} - \frac{W_D * N_D * L_D}{T_D} \quad (3)$$

Using (3) an over determined linear system with 2 variables and 20 equations was obtained. To solve this equation system, the least squared method was used. Where C is a constant and W_X is the weight of temporal segment "X" to be determined. Also, the variables MOS_X, N_X, T_X and L_X are known for each scenario. Considering the twenty test scenarios and (3), an equation linear system was obtained, which is represented by:

$$\begin{bmatrix} 1 & t_{1,2} & \cdots & t_{1,5} \\ 1 & t_{2,2} & \cdots & t_{2,5} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & t_{20,2} & \cdots & t_{20,5} \end{bmatrix} \begin{bmatrix} Ln(C) \\ W_A \\ W_B \\ W_c \\ W_D \end{bmatrix} = \begin{bmatrix} Ln(VsQM_{MOS-1}) \\ Ln(VsQM_{MOS-2}) \\ \\ \\ Ln(VsQM_{MOS-20}) \end{bmatrix} \quad (4)$$

Where: $t_{1,2}$ to $t_{1,5}$ represent the first scenario; $t_{2,2}$ to $t_{2,5}$ represent the second scenario and so on. Finally, the values of constant C and temporal segments weights were obtained.

In order to the VsQM metric proposed can be used to assess video quality in different applications, a customized player that capture the parameters, N_i, L_i and T_i and use (2) was performed. The main characteristics of this customized player are explained in the following section.

IV. Test Scenario and Application

The experimental setup consisted of a video streaming server, video client and a network emulator. The traffic between server and client was forwarded by the network emulator, which introduces artificial impairments. The experiment was carried out by streaming videos from a streaming server to a client over HTTP/TCP protocols.

In this scenario, different network impairments were configured: reduction of bandwidth and packet loss that cause impairments in the video quality, such as pauses with specific lengths. The first five video files were generated. To simplify the method, the other videos were made using a video editor, because the only video impairments were the pauses.

As previously stated, in order to monitor the buffer behavior, a customized player that captures all the events related to both buffering and playing process was used and is described as follows.

A. Customized Player

A customized player was used, in order to monitor and capture parameters or states of the buffer, used to estimate the quality of user experience when they are watching a streaming video.

The buffer states are: (a) initial buffer, which is the time period for storing minimum information to be presented at the beginning of the video. This parameter is configurable; thus, for fast networks, times can be shorter (i.e., 2 seconds), and in slow networks, longer times are recommendable (i.e., 5 seconds); (b) period of playing, video is displayed continuously without interruption and without information loss, due to the TCP data delivery reliability; (c) period of rebuffering, period of time when the buffer has no minimum information to be displayed and starts to capture data; during this period, user experiences a pause in the video. In order to clarify how customized player works, Fig. 2 shows the main buffer events.

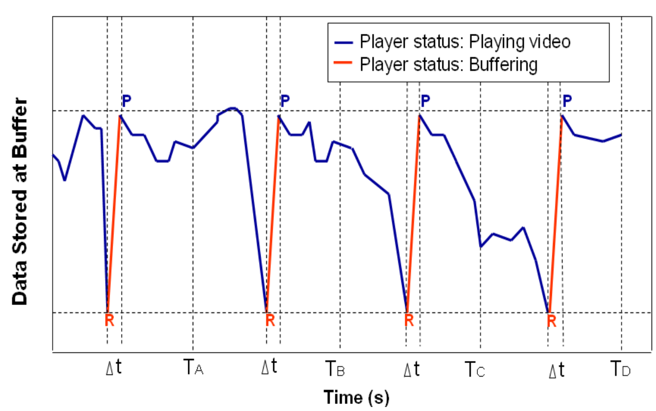

Fig. 2. Player Behavior and Events traced ("R": Rebuffering status and "P": Playing status)

With these buffer states, the following parameters could be measured: (a) number of pauses, which is the number of rebuffering events throughout the video; (b) length of each pause, which corresponds to the duration of rebuffering state; (c) frequency of pauses, which indicates the number of pauses that happened in a specific period of time; and (d) temporal location of each pause.

Temporal location parameter was considered a consequence of preliminary subjective tests, which indicated that user QoE varied depending on the temporal location of pauses in a streaming video service. For example, four pauses at the beginning of the video affects video quality in a different way from four evenly distributed pauses throughout the video or only at end of the video. In order to assess how different temporal location of pauses affects the video quality, twenty different test scenarios were built; five of these scenarios (S_1, S_2, S_3. S_4 and S_5) are depicted in Fig. 3. It is possible to see that Scenario 3 (S_3) is the same scenario presented in Fig. 2. Throughout this work, these five scenarios will be referred to in order to have a better understanding of the topics that will be treated.

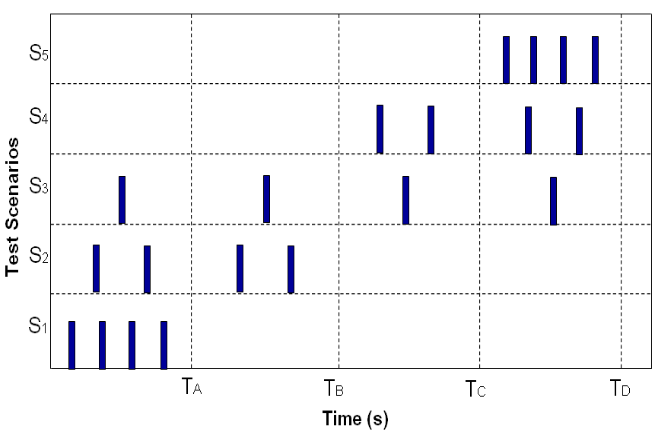

Fig. 3. Test Scenarios used as Reference

B. *Video Database*

VsQM metric was determined considering the results of the subjective video tests. For this reason, it was necessary to consider several videos with different characteristics to establish the relation of temporal segments weights with the user QoE.

Additionally to the criterion of temporal location of pauses, and the consideration that there are numerous studies that indicate that the type of video content is a determining factor in assessing the user QoE, the following video types were considered: news (reporter talking), documentary (regarding technology) and sports (soccer).

The video length was established considering that the average length of most viewed videos in current video sharing services is of 3 minutes, so the lengths of the videos used in the tests were 2 and 4 minutes. Moreover, longer videos might have discouraged the users to watch the whole video with acceptable attention, because during the tests one assessor watched ten videos on average.

With all these considerations, a database of 60 videos that considered the following criteria was created: the content type of the video, the video length, the number of pauses, the length and temporal location of each pause. Thus, for each type of video content, 20 scenarios were considered.

This set of videos was created from three original videos, the main characteristics of which are presented in Table IV.

TABLE IV
CHARACTERISTICS OF VIDEO USED IN SUBJECTIVE TESTS

Parameter	News	Documentary	Sport
Video Length (s)	240	240	240
Video/audio Format	H.264/ACC	H.264/ACC	H.264/ACC
Resolution	640x360	640x360	640x360
Frame Rate (fps)	25	29.97	29.97

As stated in ITU-T recommendation P.910, videos can be characterized using the following parameters: Temporal Information (TI) and Spatial Information (SI). Fig. 4 shows both TI and SI parameters of the three videos without pauses.

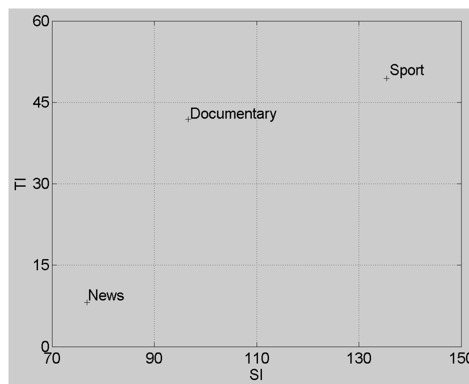

Fig. 4. Temporal Information (TI) and Spatial Information (SI) of original videos used for testing.

C. *Service Application Scenario*

As stated in the Introduction section, service providers need a feedback from end users, regarding their satisfaction level when using a specific product or service. In case of video streaming service, video quality index is an important parameter that providers can use to take different actions, in order to improve their services. Fig. 5 depicts a useful scenario, in which, VsQM metric was sent from end user to service provider using a feedback channel. Also, at the same time, this metric is shown in the users' device screens in a graphic format (e.g., 5-Stars scale).

In this work, VsQM is calculated automatically in a pre-defined time period that is configurable by the system operator. Depending on this period time, it is possible to use VsQM as input of an RDA. For non-real-time applications, the video quality metric can be used to prepare reports or to perform operations and maintain tasks. The scenario presented shows only one user, but it can be extended to a multi-user scenario. The feedback mechanism was implemented using socket interface [23].

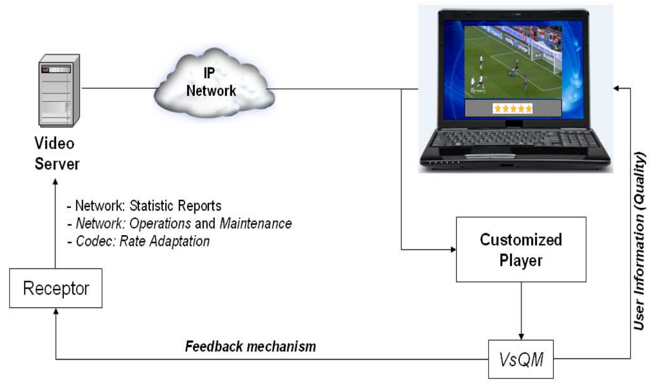

Fig. 5. Application Scenario with Feedback Mechanism for VsQM

Table V presents the algorithm performed in the player, and shows which the parameters are and the steps followed to calculate the VsQM metric.

TABLE V
ALGORITHM 1: BUFFER EVENTS AND DETERMINATION OF VSQM

Line	Statement
1	P <= instant when a playing event is triggered
2	R <= instant when a Re-buffering event is triggered
3	T_n <= Timestamp when a buffer or playing event happened
4	CT <= Current time (seconds)
5	dt <= Time period in which CT is read (Configurable)
6	τ <= Time period of each Temporal Segment (seconds)
7	k <= Number of temporal segment (In this work, k=4)
8	D <= Pause length (seconds)
9	W_i <= Weight of temporal segment *i*
10	// Events Table: (P, T and D written in vectors)
11	T_1= P <= playing started
12	T_2= R
13	T_3 = P; $D_3 = T_3 - T_2$ <= length of first pause, (written in array)
14	T_4 = R (written in array)
15	T_5= P; $D_5 = T_5 - T_4$ (written in array)
16	:
17	T_n = P or R <= Last event before CT is read
18	k =1; j=0; VsQM=0; W_A=1.3822; W_B=1.2622; W_C=1.0568; W_D=0.9875; dt=30; VsQM=0;
19	Read CT every *dt* seconds; τ = CT/4
20	for k ≤4
21	{ Case (k) {
22	k = 1 : W_i=W_A
23	k = 2 : W_i=W_B
24	k = 3 : W_i=W_C
25	k = 4 : W_i=W_D }
26	If (j*τ ≤ t < k*τ) <= t is a vector pointer
27	{ N = Number of R events (read from array)
28	D = Mean of Pause length (read from array)
29	VsQM = VsQM + (N*D*W_i) / τ }
30	k=k+1; j=j+1;
31	}
32	VsQM for sending to feedback mechanism and user device.

V. RESULTS

The total number of assessors was 96, and each video had at least fifteen scores. All assessors reported to have no vision problems and have no experience in assessing video quality tests.

As stated before, 20 different location pause and video length scenarios for each type of content video were considered, with 60 videos tested. Only the average MOS value of the three video content types for the same scenario was considered, obtaining a MOS vector length of 20. Thus, an equation linear system as presented in (4) was obtained. Finally, the values of constant C and temporal segments weights were obtained: W_A, W_B, W_C and W_D. These weights values are presented in Fig. 6.

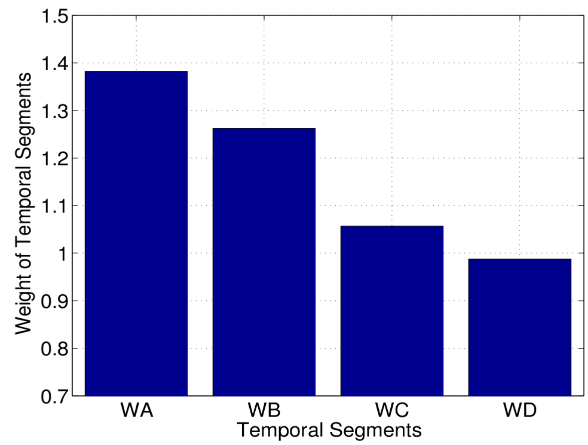

Fig. 6. Weight of Temporal Segments: W_A, W_B, W_C and W_D

As depicted in this figure, it is important to note that the initial video segment is more relevant, or it has more impairment weight, in relation to other video temporal segments. As a consequence, pauses at the beginning of the video, have a higher negative effect on user QoE, conversely to the temporal pooling methods for video quality assessment described in Section II.

These results show that the concepts of the effect of recent memory, or more recently named as working memory, is not applicable in the streaming video service with two or four-minute duration. This could be explained because the nature of the degradation is different. In the case of degradation of spatial quality, there is total loss of data. This presents some degradation such as jerkiness, blurring, brightness, mosquito noise, but the information is present. In the case of pauses, there is total loss of information that interrupts the cognitive mechanism processes.

Another important factor is that all assessors had experience in streaming video service, and unconsciously learned that, if there are disturbances at the beginning of the video, it is very likely that they will also be present throughout the video. As a consequence, they have a negative expectation from the beginning of the video.

Fig. 7 shows the relation between proposed *VsQM* metric, and both MOS index values, one of them obtained from subjective tests and the other one ($VsQM_{MOS}$) calculated by using (2). Results show that the exponential model is really reliable, because the maximum error obtained was 0.0129 at 5-point MOS scale.

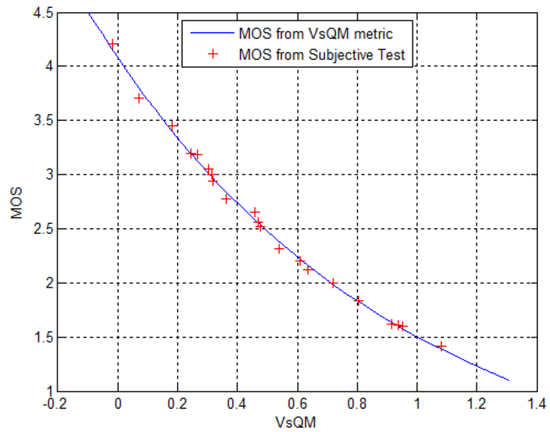

Fig. 7. Relation between Subjective MOS and $VsQM_{MOS}$

In order to determine the relation between VsQM and QoS parameters, Fig. 8 presents the relation between VsQM and Packet Loss Rate.

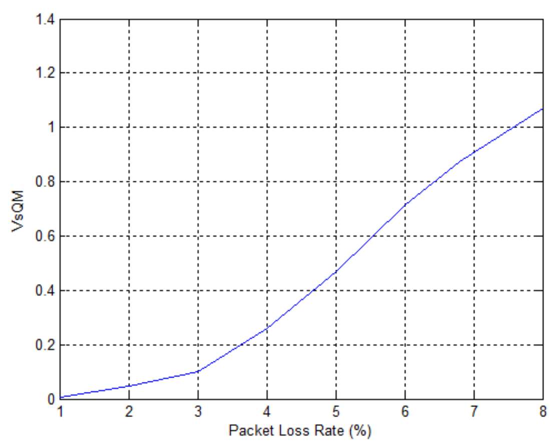

Fig. 8. Relation between Packet Loss Rate and VsQM

In Fig. 9, the relation between Pause length and objective MOS ($VsQM_{MOS}$) obtained by (2) is depicted. These five scenarios are the reference scenarios considered in Fig. 3.

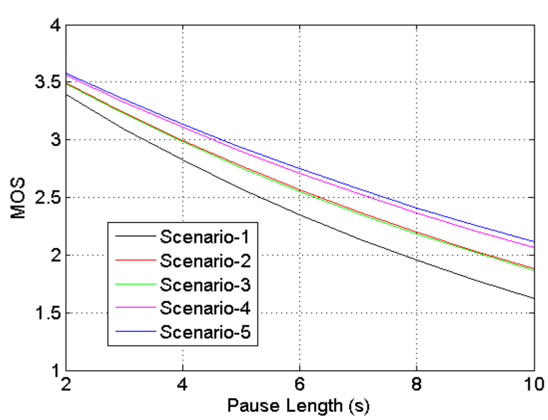

Fig. 9. Relation between $VsQM_{MOS}$ and Pause Length considering a 4-min.Video length

Fig. 10 presents the MOS index values for the same five reference scenarios, and considering video lengths of 2 and 4 minutes and pauses of 4 seconds. The goal of this graphic is to highlight the relevance of Temporal Location (*L*) parameter, in determining objective MOS using VsQM. Accordingly, both scenarios are presented. On the one hand, a VsQM model that considers *L* parameter (referred to as: "MOS (VsQM – 2min" and "MOS (VsQM – 4min"). On the other hand, *L* parameter is not considered (referred to as: "MOS (N-VsQM – 2min" and "MOS (N-VsQM – 4min").

Depending on the pause distribution scenario, the variation between the MOS index considering or not the *L* parameter can be significant; for example, in Scenario 1 with a 2-minute video length, this variation is 11.8%.

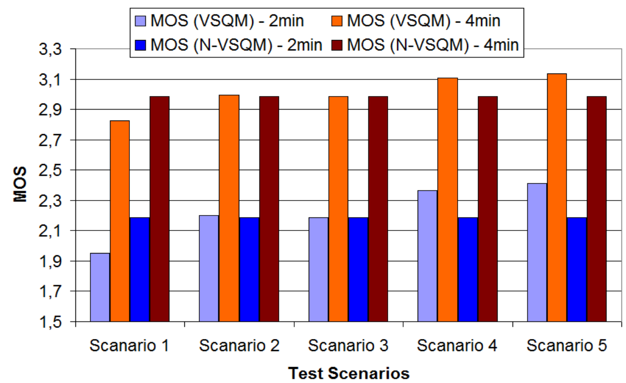

Fig. 10. $VsQM_{MOS}$ index values considering (VsQM) or not (N-VsQM) the Temporal Location Parameter using the video lengths of 2 and 4 min.

VI. Conclusion

This paper stressed the reasons why current subjective test methodologies are not appropriate to assess the user QoE of video streaming service running over TCP. As a consequence, some recommendations to conduct subjective tests were presented, which considered the variability of individual characteristics of assessors related to cognitive process mechanisms. Subjective tests, in this work, were conducted following these recommendations.

Results of subjective tests of video quality showed the relevance of considering the temporal location of pauses in a mathematical model to assess the user QoE, who is using video streaming service. It was demonstrated that pauses at the beginning of the video have a higher negative effect on user QoE in relation to the intermediate and final parts of the video. This quality degradation behavior is different from the temporal pooling methods for video quality assessment described in this paper. Furthermore, the concept of the effect of recent memory is not applicable in the streaming video service.

A new metric called VsQM was determined that takes into account the approach of temporal location parameter, and that also considers the number of pauses in a specific video segment, and the duration of each pause. Moreover, a function to map the proposed metric values into 5-point MOS scale was presented. Results show how the MOS index values, calculated from $VsQM_{MOS}$, change if the temporal locations of pauses are considered or not. It is important to note these results are based on tests with video lengths of 2 and 4 minutes.

ACKNOWLEDGMENTS

The authors thank both the Laboratory of Computer Architecture and Networks (LARC) and Signal Processing Laboratory (LPS) at Escola Politécnica - University of São Paulo for the motivation to researches in the quality of experience area in multimedia services.

REFERENCES

[1] Z. Chen, C. Lin, and X. Wei, "Enabling on-demand Internet Video Streaming Services to Multi-terminal Users in Large Scale*," IEEE* Trans. Consumer Electron., vol. 55, no. 4, pp. 1988-1996, Oct. 2009.

[2] O. Hiroki, H. Hisamatsu and H. Noborio, "Design and Evaluation of Hybrid Congestion Control Mechanism for Video Streaming," in Proceedings of 11th IEEE International Conference on Computer and Information Technology. pp. 585-590, Pafos, Cyprus, Sep. 2011.

[3] H. Hisamatsu, G. Hasegawa, and M. Murata, "Non bandwidth-intrusive video streaming over TCP," Proceedings of 8th International Conference on Information Technology, pp. 78-83, Las Vegas, U.S., Apr. 2011.

[4] L. Cai, X. Shen, J. Pan, and J.Mark, "Performance analysis of TCP friendly AIMD algorithms for multimedia applications," IEEE/ACM Transactions on Networking, pp. 339-355, Apr. 2005.

[5] H-J. Park and D-H. Har, "Subjective Image Quality Assessment based on Objective Image Quality Measurement Factors*," IEEE* Trans. Consumer Electron., vol. 57, no. 3, pp. 1176-1184, Aug. 2011.

[6] Z. Wang, A. C. Bovik, H. Sheikh, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Transactions on Image Processing, vol. 13, No. 4, Apr 2004.

[7] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Trans. Broadcast., vol. 50, no. 3, pp. 312–322, Sep. 2004.

[8] H. Kwon, H. Han, S. Lee, W. Choi, and B. Kang, "New Video Enhancement Preprocessor Using the Region-Of-Interest for the Videoconferencing*," IEEE* Trans. Consumer Electron., vol. 56, no. 4, pp. 2644-2651, Nov. 2010.

[9] J. You, A. Perkis, M. Gabbouj, and M. M. Hannuksela, "Perceptual quality assessment based on visual attention analysis," in Proc. ACM Int. Conf. Multimedia, pp. 561–564, Beijing, China, May, 2009

[10] A. K. Noorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," IEEE J. Select. Topics Signal Processing, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[11] T. Porter and X. H. Peng, "An Objective Approach to Measuring Video Playback Quality in Loss Networks using TCP," IEEE Communications Letters, vol. 15, no. 1, Jan. 2011.

[12] R. Mok, E. Chan, R. Chang, "Measuring the Quality of Experience of HTTP Video Streaming," in Proc. of IFIP/IEEE International Symposium on Integrated Network Management (IM), pp. 485-492, Dublin, Ireland, May, 2011.

[13] ITU-T Report 2007, "Definition of Quality of Experience (QoE)," International Telecommunication Union, Liaison Statement, Ref.: TD 109rev2 (PLEN/12), Jan. 2007.

[14] R. Laghari, K. Crespi, N. Molina, B. Palau, "QoE aware Service Delivery in Distributed Environment," IEEE Workshops of International Conference on Advanced Information Networking and Applications, pp. 837-842. Mar, Biopolis, Singapore, Mar. 2011.

[15] G. Nyman, J. Radun, T. Leisti, and T. Vouri, "From image fidelity to subjective quality: a hybrid qualitative/quantitative methodology for measuring subjective image quality for different image contents", in Proc. of the International Display Workshops, vol. 12, 1817-1820, 2005.

[16] G. Nyman, J. Radun, T. Leisti, J. Oja, H. Ojanen, J. Olives, T. Vuori and J. Hikkinen, "What do users really perceive: probing the subjective image quality experience", Proceedings of the IST/SPIE's: Imaging Quality and System Performance III, vol. 6059, pp. 1-7, Jan. 2006.

[17] ITU-R, "Methodology for the Subjective Assessment of the Quality of Television Pictures," Tech. Rep. BT.500-11, 2002.

[18] ITU-R, "Subjective Video Quality Assessment Methods for Multimedia Applications", Tech. Rec. P.910, Geneva, Abr. 2008.

[19] J. Blin, "New quality evaluation method suited to multimedia context: Samviq," in Proceedings of the Second International Workshop on Video Processing and Quality Metrics, VPQM'06, Arizona, U.S., Jan. 2006.

[20] J. Redi, H. Liu, H. Alers, R. Zunino, and I. Heynderickx, "Comparing Subjective Image Quality Measurement Methods for the Creation of Public Databases," in Proc. of SPIE-IS&T Electronic Imaging, SPIE vol. 1, no. 7529, California, U.S., Jan. 2010.

[21] K. Seshadrinathan, A.C. Bovik, "Temporal hysteresis model of time varying subjective video quality," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1153-1156, Jul. 2011.

[22] J. You; J. Korhonen, J.; A. Perkis, "Attention modeling for video quality assessment: Balancing global quality and local quality," in Proc. of IEEE International Conference on Multimedia and Expo (ICME), pp. 914-919, Singapore, Jul. 2010.

[23] D. Rodríguez, M. Arjona, "VoIP Quality Improvement with a Rate-determination Algorithm", in Proc. of International Workshop on Telecommunications, pp. 30-34, Sao Paulo, Brazil, Feb. 2009.

BIOGRAPHIES

Demóstenes Zegarra Rodríguez received his B.S. degree in Electronic Engineering from the Pontifical Catholic University of Peru and the M.S. degree from University of São Paulo - USP (2009). Nowadays, he is a Ph.D. student in Electronic Systems at USP, with solid knowledge in Telecommunication Systems and Computer Science based on 12 years of professional experience in important companies. His current interest includes, QoS and OoE in Multimedia services, Digital TV and architect solutions in Telecommunication Systems.

Julia Issy Abrahão got her PhD degree in Ergonomics from the Conservatoire National des Arts et Metiers (1986) in Paris. Her Post-doc was obtained from École Pratique des Hautes Études and Université Paris V. Nowadays she is an associate researcher at Universidade de Brasília and a visiting professor at Escola Politécnica of the Universidade de São Paulo. Her current research interests involve production engineering and architecture with emphasis on ergonomics. She is also involved with research into cognitive ergonomics, usability, automation, ergonomic analysis of working activities, new technologies and health.

Dante Coaquira Begazo received his B.S. degree in Electronic Engineering from the National University of San Agustín of Arequipa (UNSA), and the M.S. degree from University of São Paulo - USP (2011). Nowadays, he is PhD student at USP. He has experience in IT companies. His current interest includes the areas of evaluation of video quality on IP networks, video processing, and audio-visual speech recognition.

Renata Lopes Rosa received her B.S. degree in Science Computer from UNIFEI, Brazil, and the M.S. degree from University of São Paulo - USP (2009). She is a Ph.D. student at Escola Politécnica of the University of Sao Paulo (EPUSP). Her current research interest includes computer networks, quality of experience of multimedia service, social networks and recommendation systems. Nowadays, she is a Professor at SENAC University.

Graça Bressan was granted her PhD in Electronic Engineering (1986) by Escola Politécnica of the University of São Paulo (EPUSP). Her current research interests include computer networks and digital television focusing on aspects of distributed systems, distributed middleware, QoS mechanisms, collaborative virtual environment, middleware for Digital TV, interactive digital TV, video-conferencing, modeling and performance analysis of networks, and applications in distance education.