

# Video Quality Assessment in Video Streaming Services Considering User Preference for Video Content

Demóstenes Z. Rodríguez, Renata L. Rosa, Eduardo A. Costa, Julia Abrahão, and Graça Bressan

**Abstract** — *In video streaming service, the user's Quality of Experience (QoE) is not only related to video signal quality received at consumer's devices, the users' subjectivity must also be considered. In this context, a video quality assessment method that takes into account the user's preference for video content is proposed in this research. In order to perform this task, the users' profiles that include their preferences were stored in a video server. Then, subjective tests of video quality assessment were conducted, in which evaluators had different video content preferences. Results show that the evaluators' QoE is highly correlated with the user's preference for video content type. Based on these experimental results, a function named Preference Factor (PF) is defined and used to adjust the quality index values obtained by an objective video quality metric running in the end user's device. The PF function also depends on video content type and quality index score. Using the PF function, the enhanced Video streaming Quality Metric (e-VsQM) is proposed and the results of its performance evaluation demonstrate that PF improves an objective video quality metric. Furthermore, e-VsQM has low complexity and can be utilized in different video services. Thus, an application scenario is presented, in which the proposed video quality metric is implemented<sup>1</sup>.*

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

## I. INTRODUCTION

In recent years, many video service applications over the Internet have gained popularity generating a huge volume of traffic every day, in which one of the most successful video services is video streaming. In order to meet end users' expectations, researches on video QoE are relevant for users, content providers, and consumer electronics manufacturers.

In video streaming services, the user's QoE assessment is a complex task because QoE can be influenced by different factors, which can be classified in the following categories, human, system and context [1]. Human factors are concerned with the user's subjectivity composed of different mechanisms of human information processing, from sensorial to cognitive processes that include personal characteristics, such as users'

preferences [2]-[4]. System influence factor is strongly related to end users' devices. In multimedia services, design characteristics and processing capacities of electronic devices impact on user's QoE [5], [6]. Context influence factors are concerned with the space and time in which a service is used. Also, social and economic aspects are considered.

The multimedia consumer electronics involves video adaptation techniques to deliver a good video quality service [7] with an optimized use of network bandwidth, and computational resources of devices to improve user's QoE [8]. Also, researches in video encoding solutions start to use video content identification to outperform algorithms with popular segmentation techniques, improving both network and device performance [9].

Subjective video quality assessment methods are necessary to quantify user's QoE of multimedia services. As a result, an index value, mostly known as Mean Opinion Score (MOS), is obtained, which is the mean of the scores granted by at least fifteen evaluators. Furthermore, device manufactures improve their quality services taking into account subjective test results [10]. In turn, objective video quality metrics use a mathematical model or an algorithm to establish a video quality score, which reflects the experimental results of subjective video quality assessment. Thus, an objective video quality metric is more reliable if its scores are highly correlated with the subjective assessment provided by evaluators [11]. Analyzing the state of the art, many proposals of objective video quality metrics are verified; each one with different approaches, but none of them considers the user's subjectivity characteristics as an input parameter.

In this context, the major contribution of this research is to demonstrate that objective video quality metrics, running on users devices, can be improved by considering the user's preferences for video content type; therefore, video quality metrics can reflect a more realistic user's QoE.

This paper is an extended contribution to [12], and its main goal is to propose a new objective video quality metric, named e-VsQM, in the video streaming service that considers the user preference for video content in addition to the video signal quality. In order to define the e-VsQM metric, subjective tests were conducted, in which assessors with different preferences for video content type participated. As a result, a function that works as a correction factor was obtained for each video content category. Once the proposed metrics has been defined, a network architecture solution is introduced, in which the e-VsQM metric is implemented.

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

D. Z. Rodríguez, R. L. Rosa, J. Abrahão, and G. Bressan are with the School of Engineering - University of São Paulo, Brazil. (e-mails: {demostenes, gbresan}@larc.usp.br, {jabrahao, rrosa}@usp.br).

E. A Costa is with the University of Brescia, Italy (e-mail: e.costa@unibs.it).

It is worth noting that using the proposed solution, the video service providers will have personalized information regarding the user's QoE. Hence, this information can be useful to offer additional features to specific users, or also be included in solutions such as (DASH) [13]; for instance, users with no preference for a specific video content are not interested in obtaining a higher video resolution, avoiding unnecessary resources of both network and user's device.

The remainder of this paper is structured as follows. Section II presents a review of Video Quality Assessment Methods and QoE definition and its influence factors. Section III introduces the proposed Video Quality Model used to determine the e-VsQM. Section IV shows the Test Implementation and the proposed Network Architecture. Section V presents the results that include the e-VsQM performance evaluation. Finally, section VI presents the conclusions.

## II. VIDEO QUALITY ASSESSMENT METHODS AND QUALITY OF EXPERIENCE

In this section, firstly, the most popular current test methodologies for video quality assessment are reviewed and classified. Secondly, some considerations to determine the QoE of multimedia services are treated.

### A. Current Methodologies for assessing Video Quality

Currently, there are many image and video quality assessment methods, each meeting different purposes. These methods can be classified in different ways depending on the criteria set adopted, as verified in Fig. 1.

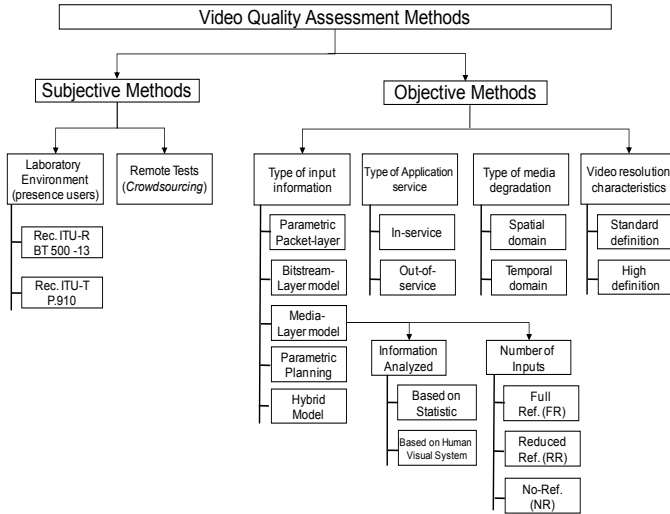


Fig. 1. Classification of Video Quality Assessment Methods using different criterions.

As can be observed in Fig. 1, in general, video quality assessment methods can be performed by using subjective or objective test methods.

There are several subjective tests methods for video quality assessment, most of them described in the ITU recommendations ITU-R BT-500 [14] and ITU-T P.910 [15]. The latter method is focused on multimedia services. These

tests are generally conducted under laboratory conditions, in which the supervisor explains the test instructions to the assessors. Later, assessors watch a test video; they grant an adjective score using different scales. One of the most accepted scales is the 5-point MOS scale described in the Absolute Category Rating (ACR) method standardized in ITU-T Recommendation P.910 shown in Table I.

TABLE I  
ITU-T 5-POINT SCALE – ABSOLUTE CATEGORY RATING

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

Nevertheless, in recent years, some researches [16]-[19] have demonstrated the possibility to perform quality assessment of multimedia services through remote assessors using the Internet. The main purpose of this new methodology is to considerably reduce the time to perform subjective tests, because, this is one of the main difficulties in conducting subjective tests in laboratory environments.

In order to perform remote tests, both commercial platforms [16] and social networks [20] can be used. In the first one, assessors receive a monetary compensation for each evaluation. In the second one, assessors perform the same task without any payment.

Objective methods for video quality assessment can be classified by using several considerations. They are classified depending on the type of input data into the following five categories [21]: (1) Parametric packet-layer models try to predict the MOS index only from the packet-header information and do not consider the media signals. (2) Bitstream-layer models use encode bitstream information and also packet-layer information. (3) Media-layer models use the video signal to estimate the MOS index. (4) Parametric planning models require a priori knowledge about the system being tested. (5) Hybrid models combine two or more of the preceding models.

Depending on the type of application service, objective methods are divided into two categories [11]: (1) In-service methods have time constraints because they are used in real time applications [22], [23]. (2) Out-of-service methods do not have time restrictions, and are used in different tasks, such as, video codec performance evaluation and video streaming services [24], [25].

Considering video transmission services, different transport layer protocols can be used, the most common being Transport Control Protocol (TCP) and User Datagram Protocol (UDP). Thus, different types of video impairment are obtained. TCP protocol is a reliable protocol because it guarantees the delivery of all packets to end user. But, if there are packet losses in the transmission channel, 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 pauses occur. Because of the impairment nature, the video quality metrics that assess this type of impairment are classified into the temporal domain category [26]-[28]. Conversely, UDP does not guarantee the packet delivery; then, in case of packet losses, each video frame arriving at the end point will present lack of information, considering the spatial domain [29].

Recently, novel models have been proposed for assessing high definition video quality (HD) [30]-[32]. Also, the Video Quality Expert Group (VQEG) proposed validation tests for objective quality metrics applicable to HD videos [33].

As illustrated in Fig. 1, the category of media-layer model is sub-classified according to the information analyzed. The well known objective metrics, Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are calculated based on statistical analysis of the pixels information. In turn, the metrics Structural Similarity (SSIM) [34], Video Quality Metric (VQM) [35] and algorithms based on Region of Interest (RoI) [36] or attentions maps [37], [38] are based on the Human Visual System.

Finally, the category of media-layer model is also sub-classified according to the number of inputs used for a video quality metric [39].

After this general review regarding objective video quality metrics, it can be concluded that these metrics are focused on determining the human visual quality perception. In order to determine the users' QoE, other information regarding their subjectivity needs to be considered in the objective metric model, such as, preference for video content.

### *B. Quality of Experience and its Influence Factors*

In communication services, the notion of quality has been associated to the concept of Quality of Service (QoS) for many years. However, QoS is related to technical aspects and does not treat the user's subjectivity. The concept of QoE appeared to fill this lack and its definition most frequently used, given by ITU-T is: "the overall acceptability of an application or service, as perceived subjectively by the end user" [40].

Recently, the concept of QoE is extended to: "the degree of delight or annoyance of a person whose experience involves an application, service, or system. It results from the person's evaluation of the fulfillment of his or her expectations and needs with respect to the utility and/or enjoyment in the light of the person's context, personality and current state" [41]. Considering this definition, one can infer that there are different factors influencing the QoE. In [1], these influence factors are classified into three categories, system, context and human.

System influence factors refer to properties and characteristics that determine the technically produced quality of an application or service, and they are sub-classified into media-related, network-related and device-related system influence factors [42], [43]. Hence, in video streaming service, the video signal quality assessed at the end user's device influences QoE, in which the signal quality is related to media

and network factors.

Context influence factors are related to, physical, temporal, social and economic contexts. Physical and temporal contexts describe the characteristics of location and space, including acoustic and lighting conditions, and also time of day, week, month or season. The social context is defined by the interpersonal relations existing during the experience. Finally, the economic context considers costs, subscription type, or brand of the system or application [1].

A human influence factor is any characteristic of a human user. The characteristic can describe the demographic and socio-economic back-ground, the physical and mental constitution, or the user's emotional state [1], [44]. Preferences and attitudes are considered as factors that may influence QoE at higher level of human cognitive process. Thus, human influence factors are related with the user's subjectivity, for instance, in video streaming service, the user's preference for a video content type is a factor that influences QoE.

In short, user's QoE can be influenced by a wide range of factors, which are complex and interrelated. Hence, to determine the user's QoE in video streaming service, the video signal quality perception at end user's devices needs to be complemented with other criteria set related with sensorial processing, human cognitive process and psychological approaches. Consequently, QoE needs to be a multi-disciplinary study, and incorporate disciplines such as: cognitive science, psychology, sociology, economy, and information technology.

In order to better analyze subjective test results, the assessors' profiles need to be considered. Some information such as, age, gender, education level, personal knowledge, attitudes, expectations, emotions and preferences need to be known [2]. In this research, the user's preference for video content is taken into account to improve an objective video quality metric.

## **III. THE PROPOSED VIDEO QUALITY METRIC**

In order understand the impact of users' preferences on visual QoE, preliminary video quality subjective tests were performed, in which the users' profiles were considered, specifically their explicit preference for video content type. The experimental results of these subjective tests are presented in Fig. 2 [12].

These results clearly showed that evaluators preferring a specific video content granted different score compared with evaluators without preference for the same content type. Then, the conclusion was that an objective video quality metric has to consider the user's preference for video content type and an extensive number of subjective tests were conducted.

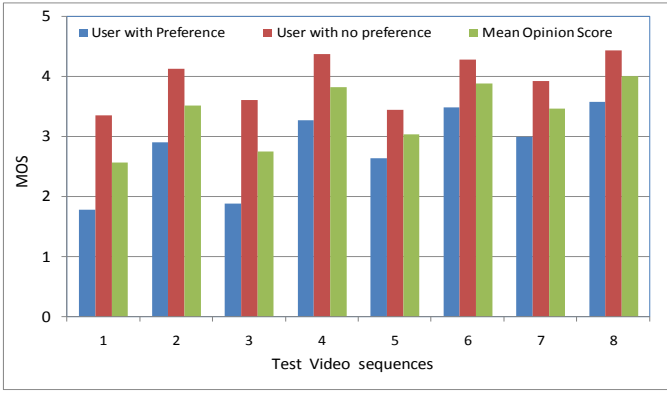


Fig. 2. Subjective Test Results of eight video sequences.

As stated before, current objective video quality metrics do not consider the user's preference, because their algorithms are only based on video signal, network or application parameters.

Hence, the concept of Preference Factor (PF) is introduced herein. PF is related to the human factor and it is a function that works as a correction factor, because it adjusts the MOS index scores obtained by an objective metric, in order to improve the correlation with the real user's QoE.

Experimental results of subjective tests showed that PF values depend on:

- User's preference, which could be, preference ( $p$ ) or no preference ( $np$ );
- Video content type; this work considered three content types, sport, documentary and news;
- Score level of MOS index values.

In this work, twenty different impairment videos for each content type were evaluated, and are explained in the next section. An MOS index for each video assessed resulted from the subjective tests, and these MOS scores will be used to find the ratios between the scores given by the users with and without preference in relation to the global MOS index.

Variables  $Rp_i$  and  $Rnp_i$  are defined and represent the ratios between MOS scores granted by users with and without preference, respectively. They are presented in (1) and (2).

$$Rp_i = \frac{MOS_i^{preference}}{MOS_i^{mean}} \quad (1)$$

$$Rnp_i = \frac{MOS_i^{no-preference}}{MOS_i^{mean}} \quad (2)$$

In which  $MOS_i^{mean}$  is the mean value considering the total number of users with and without preference in conjunction for the video test number " $i$ ". In this work the maximum value of " $i$ " is twenty. Because the number of assessors with preference and no-preference in each video assessed " $i$ " was the same, the relation between  $Rp_i$  and  $Rnp_i$  is given by:

$$Rnp_i = 2 - Rp_i \quad (3)$$

For each video content type, there are twenty relations of  $(MOS_i^{mean}, Rp_i)$  or  $(MOS_i^{mean}, Rnp_i)$  obtained from the subjective tests. With this information, a function named  $PF_p^{CT}$  that represents the PF function based on the  $Rp_i$  values for each video content type can be modeled empirically by:

$$PF_p^{CT} = \alpha * \ln(MOS_p^{mean}) + \beta \quad (4)$$

Where the  $CT$  index represents the video content type and is limited to Sport, Documentary and News, and the  $p$  index represents that the user prefers  $CT$ . Also,  $MOS_p^{mean}$  values are in the range from 0.5 to 4.5. Table II shows the values for each variable that can well fit the function presented in (4).

TABLE II  
VARIABLE VALUES FOR THE FUNCTION  $PF_p^{CT}$  CONSIDERING DIFFERENT CONTENT TYPES

Video Content Type (CT)	$\alpha$	$\beta$
Sport	0.741	0.219
Documentary	0.466	0.284
News	0.452	0.352

It is worth noting that results show that the logarithmical model is really reliable, because the maximum error obtained by using  $\alpha$  and  $\beta$  values in (4) for Sport, Documentary and News were 0.02, 0.03 and 0.03, respectively at a 5-point MOS scale.

Similarly, to determine a function that represents the  $Rnp_i$  values called  $PF_{np}^{CT}$ , equation (3) and (4) are used, as presented in (5).

$$PF_{np}^{CT} = 2 - \alpha * \ln(MOS_p^{mean}) - \beta \quad (5)$$

Where,  $\alpha$  and  $\beta$  assume the values presented in Table II.

Once the  $PF_p^{CT}$  and  $PF_{np}^{CT}$  functions are defined, they can be used in different video streaming services, in which the  $MOS_p^{mean}$  variable is replaced in (4) and (5) for the MOS index obtained by an objective metric as shown in Fig. 3. Some outputs of objective metrics are not mapped to the 5-point MOS scale, making a scale conversion necessary.

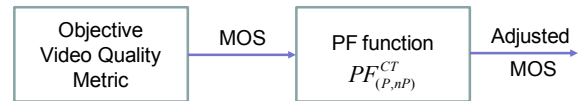


Fig. 3. Preference factor function used to adjust a MOS index obtained by an objective video quality metric.

The video quality metric named e-VsQM is proposed herein, and is an improvement of the objective metric named VsQM, defined in a previous work [28]. The VsQM is used to assess video streaming services that use the TCP as the transport protocol layer. Thus, the VsQM metric uses the information captured by a customized video player, and is

based on the number, duration and temporal location of pauses that occur during a video streaming transmission. The VsQM metric is mapped to the 5-point MOS scale and is defined by

$$VsQM = C * \exp\left(-\sum_{i=1}^k \frac{N_i * L_i * W_i}{T_i}\right) \quad (6)$$

Where:

- $C$  is a constant for scaling purposes;
  - $N_i$  is the number of pauses;
  - $L_i$  is the average length of pauses, in seconds, which happened in the same temporal segment;
  - $W_i$  is a weight 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.

Finally, e-VsQM metric improves the VsQM metric performance by adding the functions defined in (4) or (5), both represented by  $PF_{(P,nP)}^{CT}$  as shown in (7).

$$e-VsQM = VsQM * PF_{(P,nP)}^{CT} \quad (7)$$

In order to use the e-VsQM metric, the user's preference for video content needs to be stored in the video server application. Also, each video file needs to be previously classified into one of the content-type categories.

#### IV. TEST IMPLEMENTATION AND THE PROPOSED NETWORK ARCHITECTURE

In this section, firstly, the video database used as test material in the video quality subjective test is explained. Secondly, a service application scenario is presented, in which the e-VsQM is implemented. And finally, the crowdsourcing method used to conduct remote subjective tests is treated.

##### A. Video Database

It is worth noting that video databases such as LIVE [45] and VQEG [33] only consider video lengths of 10 seconds, and users' profiles are not presented. Using this video length, the effects of degradation caused by pauses cannot be properly evaluated. In this work, the video length was established considering that the average length of the most viewed videos in current video sharing services is in the range of 3 to 5 minutes; the lengths of the videos used in the laboratory tests was thus 4 minutes. Also, videos with different characteristics had to be considered to establish the relation of video content type with the user's preferences. The following video content categories were considered: news (reporter talking), documentary (regarding technology) and sports (soccer).

Based on all these considerations, a database of 60 videos considering 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 video content type category, 20 impairment videos were considered. For clarification purposes, Fig. 4 shows four impairment videos with different number of pauses and each pause with different lengths and temporal locations.

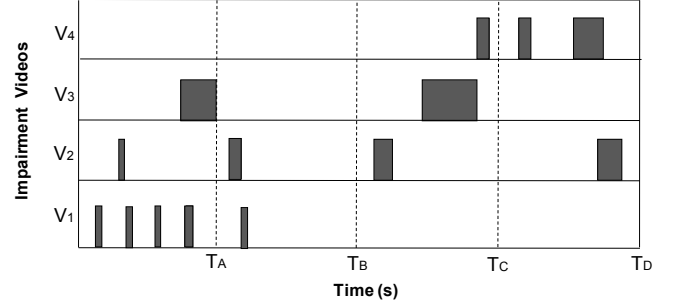


Fig. 4. Example of four impairment videos created inserting pauses with different lengths and temporal locations

Also, as stated in ITU-T recommendation P.910, videos can be characterized using the following parameters: Temporal Information (TI) and Spatial Information (SI). The set of impairment videos was created from three original videos, the main characteristics of which are presented in Table III.

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)	30	30	30
Temporal Information (TI)	8	43	51
Spatial Information (SI)	76	96	134

##### B. Service Application Scenario

The test scenario considers a video streaming service over HTTP/TCP. In order to insert some network impairments, a network emulator was used as a transmission channel. Thus, number of pauses and their lengths are controlled reducing the bandwidth of the transmission channel. This procedure was performed during preliminary tests. However, to guarantee that pauses appear at the same time instant and with a fixed duration in all the videos assessed, these were edited. Then, during subjective tests, impairment videos stored at video server were transmitted without any network degradations.

In the client device, the VsQM metric is implemented. For this, a customized player was used, in order to monitor and to capture the following buffer parameters: (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 the rebuffering state; and (c) temporal location of each pause. With these parameters and using (6), the VsQM value is determined. This information is displayed at the device screen in graphic format, and it is also sent to the video server using a notification mechanism.

Fig. 5 presents the proposed network architecture with emphasis on the user's device implementation. In this work,



the notification mechanism was implemented using a socket interface [46].

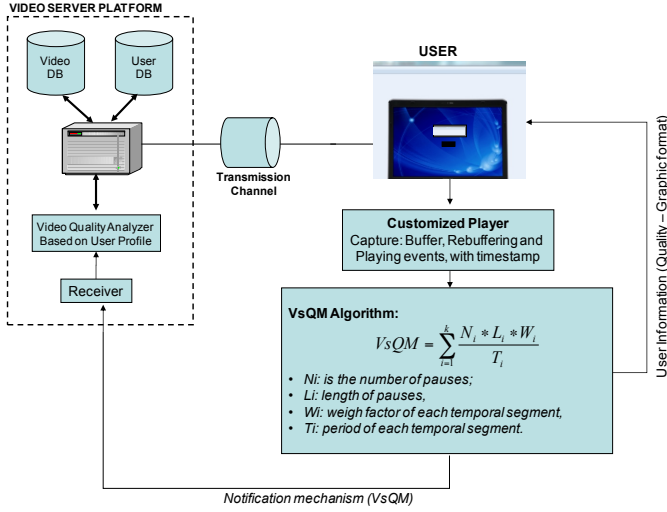


Fig. 5. Network architecture of video streaming service with emphasis on the video quality metric implemented at the user's device.

The processing resource required for VsQM implementation is much lower than video decoding processing and does not affect the performance of the current electronic devices. Also, the client application has an interface in which users can register their profiles including their preferences for video content. This information is storage in the video server.

Fig. 6 shows the implementation in the video server platform, in which the users' profile and the classified video content types were previously stored. The VsQM value is received from the client and used with the video identification ( $V_{id}$ ) and user's preference as inputs of the algorithm that calculates the e-VsQM value. The e-VsQM algorithm has low complexity, and is presented in Table IV.

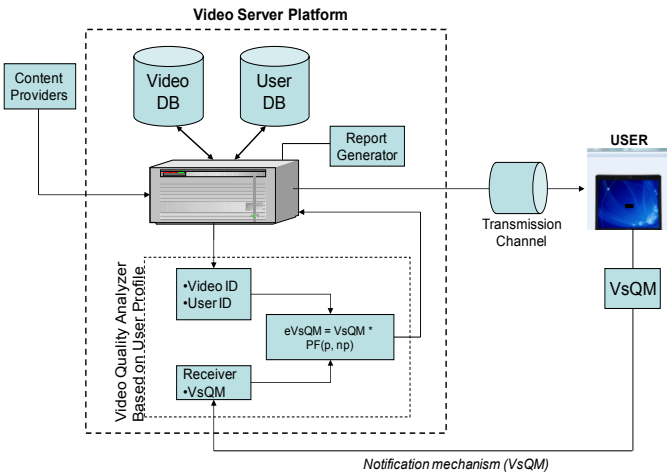


Fig. 6. Network architecture of the video streaming service with emphasis on the solution implemented in the video server.

TABLE IV  
ALGORITHM 1: VIDEO QUALITY METRIC AWARE OF USER PREFERENCE FOR VIDEO CONTENT

Line	Statement
1	Video ID: $V_{id}$
2	$CT_1$ : sport = $\{V_{s1}, V_{s2}, V_{s3}, \dots, V_{si}\}$ // i.e.: $V_{id} = V_{s1}$
3	$CT_2$ : documentary $\{V_{d1}, V_{d2}, V_{d3}, \dots, V_{dj}\}$
4	$CT_3$ : news $\{V_{n1}, V_{n2}, V_{n3}, \dots, V_{nk}\}$
5	$Up$ = User preference for video CT
6	// Example: $Up = \{CT_1, CT_2\}$
7	Read ( $VsQM$ )
8	If ( $Vid \in Up$ )
9	$eVsQM = VsQM * PF_p^{CT}$
10	Else
11	$eVsQM = VsQM * PF_{np}^{CT}$

### C. Subjective Tests Using the Crowdsourcing Method

In order to evaluate the e-VsQM performance, additional tests were conducted by remote assessors connected to a commercial platform, using the crowdsourcing method briefly described in section II. To perform this task, a web interface was built and is presented in Fig. 7. This interface contains the video files and the instructions to perform the video quality assessment tests. The assessors' tasks are basically:

- First, registering in the crowdsourcing platform as a worker;
- Filling in the user's profile questions, such as age, gender, geographic location, and mainly the user's preference for video content;
- Downloading and assessing the video file.
- Answering some test validation questions, so as to validate if the assessor watched the totality of the video sequence. If assessors answer incorrectly, their results are not considered and they do not receive a monetary compensation. In this case, another worker will perform the same task.

The screenshot shows a web interface for a 'Video Mean Opinion Score Test - MOS - Lab Poli - USP'. It includes a language selector (English, Portuguese, Spanish), a table of MOS scale ratings (1 to 5) with definitions and descriptions, and a section for 'Instructions for testing'. Below this, there are 'Questions to determine users profile' (age, preference on video content) and a section for 'Video files to be downloaded' (Video-Condition 1, 2, 3) with 'select one option' dropdowns. A 'Submit' button is at the bottom.

Fig. 7. Interface for video quality assessment using the crowdsourcing method.

## V. RESULTS

In a first phase, subjective tests were performed in a laboratory environment, and their results served to determine the  $PF_p^{CT}$  and  $PF_{np}^{CT}$  functions. In a second phase, the performance of the e-VsQM is evaluated using the crowdsourcing method. These phases are explained as follows.

### A. Subjective Test in a Laboratory Environment

A 21.5-inch LCD monitor was employed with the following characteristics: 1920x1080 pixels resolution, widescreen ratio of 16:9. The test environment had no reflecting ceiling walls or floors, or any disturbing objects. The tests were conducted in 13 weeks, and during this period the same test room was kept constant. All the tests were performed individually and a time limit was not enforced. An instruction session was performed before the tests, in which the assessors were shown sample videos and the experiment process was explained. In the tests, an observation distance of 50 cm to 60 cm was considered, and assessors used the scale presented in Table I. Each video received at least 15 scores and the total number of assessors was 107. With the test results, a statistical analysis was performed and no assessor was identified as an outlier.

The  $PF_p^{CT}$  and  $PF_{np}^{CT}$  functions defined by (4) and (5) are depicted in Fig. 8 and Fig. 9, respectively.

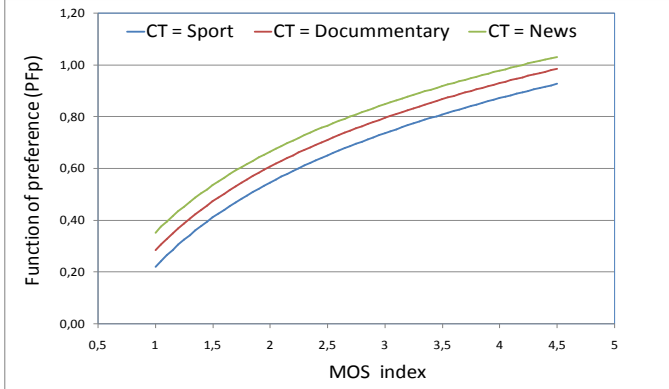


Fig. 8. Relation between MOS index and the Preference Factor function ( $PF_p^{CT}$ ) considering sport, documentary and news as content types.

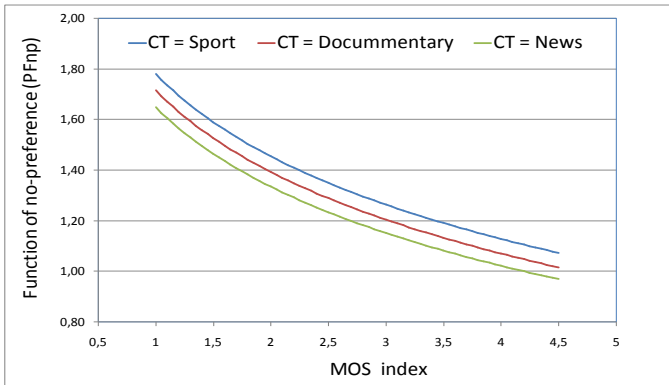


Fig. 9. Relation between MOS index and the Preference Factor function ( $PF_{np}^{CT}$ ) considering sport, documentary and news as content types.

Functions  $PF_p^{CT}$  and  $PF_{np}^{CT}$  are observed to change according to both the MOS index values and the video content type. Analyzing the  $PF_p^{CT}$ , the sport video content type has the lowest values, and news presents the highest values, which means that the user's QoE is more degraded for users preferring sports videos. Conversely, if a user does not prefer sports, the QoE is the least affected, because the  $PF_{np}^{CT}$  function has the highest values for sports. Additionally, the increasing or decreasing rate of the functions  $PF_p^{CT}$  and  $PF_{np}^{CT}$  changes according to the MOS values.

### B. Performance Evaluation of the e-VsQM

The performance evaluation of the e-VsQM metric was performed using the crowdsourcing method. For this, additional 12 impairment videos were used as test material, in which 4 videos for each content type were considered. In order to have different impairment videos in relation to that used in the laboratory environment, a different number and distributions of pauses were inserted. Also, the video length was 6 minutes.

The crowdsourcing platform rules stipulate that each work requirement, named campaign, should be evaluated by at least 30 users. Thus, for each rated video, 3 campaigns were sent, achieving 90 MOS scores for each video assessed.

The first analysis meant to determine the reliability of the results obtained by remote users regarding the results from the laboratory environment. Six video sequences were used. The results of both remote and laboratory tests are shown in Table V. With these results, the Pearson correlation coefficient was 0.9797, which gives enough confidence to perform additional tests using the same method.

TABLE V  
RESULTS OF VIDEO QUALITY TESTS PERFORMED IN A LABORATORY ENVIRONMENT AND USING CROWDSOURCING METHOD

Video content category	Video sequence	Laboratory	Crowdsourcing
Sport	1	3,25	3,28
Sport	2	2,87	2,94
Documentary	3	3,52	3,63
Documentary	4	2,98	2,87
News	5	3,72	3,79
News	6	3,22	3,34

Fig. 10 presents the MOS values obtained by the remote users (crowdsourcing), and the objective MOS values estimated by the VsQM and e-VsQM using (6) and (7), respectively. Each assessor only evaluates video content types of his or her preference. Four impairment videos were considered. Video sequences indexed with "1" and "4" represent videos with the lowest and highest number of pauses, respectively.

As depicted in this figure, VsQM presents a sole MOS score for video sequences with the same index, for instance, "Esp.1", "Doc.1" and "News1", because VsQM does not consider the user's preference for video content. Using the functions  $PF_p^{CT}$  and  $PF_{np}^{CT}$ , e-VsQM obtains MOS scores highly correlated with Crowdsourcing results. The Pearson

correlation coefficients obtained are  $\rho_1(e\text{-VsQM}, \text{Crowdsourcing}) = 0.99$  e  $\rho_2(\text{VsQM}, \text{Crowdsourcing}) = 0.97$ . The maximum errors obtained at a 5-point MOS scale were 0.2 and 0.8 for e-VsQM and VsQM, respectively. Also, the sports videos were observed to obtain the lowest MOS score values if only the video sequences with the same index are considered.

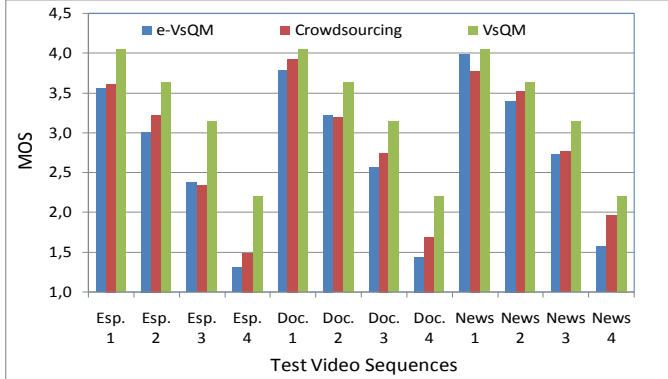


Fig. 10. Performance comparison of the objective MOS determined by e-VsQM and VsQM in relation to Crowdsourcing results.

## VI. CONCLUSION

Experimental results highlight the relevance of considering the assessor's preference for video content in video quality assessment tests. Thus, assessors with an explicit preference granted the lowest MOS scores. Based on these results, functions  $PF_p^{CT}$  and  $PF_{np}^{CT}$  were defined, which depend on user's preference, video content category and MOS scores. These functions work as a correction factor and meant to consider the user's subjectivity in an objective metric. These functions were used to improve the VsQM metric [28]. As a result, a new no-reference objective metric for video streaming service running over TCP named e-VsQM was defined.

The results from the e-VsQM performance evaluation demonstrated that functions  $PF_p^{CT}$  and  $PF_{np}^{CT}$  can be used to improve the accuracy of an objective video quality metric to estimate the user's QoE, because they are related to the human factors. It is important to note that e-VsQM considers human and system influence factors of QoE discussed in section II.

Additionally, a network architecture solution is presented, in which applications to calculate the e-VsQM are installed in both video server and user's device. Hence, video service providers can have more realistic and personalized information regarding the user's QoE, which can be useful to new applications. Also, e-VsQM has low complexity and could be utilized in different video services.

## ACKNOWLEDGMENTS

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

## REFERENCES

- [1] U. Reiter et al. "Factors influencing Quality of Experience," in *Quality of Experience*, S. Moeller and A. Raake, Eds. London: Springer, 2014, pp. 55-72.
- [2] B. Dedtweiler-Bedell, J. Dedtweiler, and P. Salovey, "Mood-Congruent Perceptions of Success Depend on Self-other Framing," *Psychology Press-Cognition and Emotion*, vol.20, no. 2, pp. 196-216, 2006.
- [3] M. Silva, J. Groeger, and M. Bradshaw, "Attention-memory interactions in scene perception," *Spatial Vision*, vol. 19, no.1, pp. 9-19, Netherlands, 2006.
- [4] D. Rodriguez, J. Abrahão, D. Begazo, R. Lopes, and G. Bressan, "Video quality subjective assessment considering cognitive criteria and user preferences on video content," in *Proc. 18th ACM Brazilian Symposium on Multimedia and the Web*, pp. 269-272, Sao Paulo, Brazil, Oct. 2012.
- [5] A. Khan, L. Sun, J. Fajardo, F. Liberal, E. Ifeachor, "Impact of end devices on subjective video quality assessment for QCIF video sequences," in *Proc. International Workshop on Quality Multimedia Experience*, pp. 177-182, Mechelen, Belgium., Sep. 2011.
- [6] A. Floris, L. Atzori, G. Ginesu, and D. Giusto, "QoE assessment of multimedia video consumption on tablet devices," in *Proc. Globecom Workshops: Quality of Experience for Multimedia Communications*, pp. 1329-1334, California, US., Dec. 2012.
- [7] F. Pereira, "Sensations, perceptions and emotions: towards quality of experience evaluation for consumer electronics video adaptations" in *Proc. International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, US., Jan. 2005.
- [8] A. Banitalebi-Dehkordi, M. Pourazad, and P. Nasiopoulos, "Effect of high frame rates on 3D video quality of experience," in *Proc. IEEE International Conference on Consumer Electronics*, pp.416-417, Las Vegas, US., Jan. 2014.
- [9] V. Adzic, H. Kalva, and B. Furht, "Optimizing video encoding for adaptive streaming over HTTP," *IEEE Trans. on Consumer Electron*, vol.58, no.2, pp.397-403, May 2012.
- [10] H-J. Park and D-H. Har, "Subjective Image Quality Assessment based on Objective Image Quality Measurement Factors," *IEEE Trans. on Consumer Electron*, vol. 57, no. 3, pp. 1176-1184, Aug. 2011.
- [11] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison.," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 165-182, Jun. 2011.
- [12] D. Rodriguez, R. Lopes, and G. Bressan, "Video quality assessment in video streaming services considering user preference for video content," in *Proc. International Conference on Consumer Electronics*, pp. 570-571, Las Vegas, US., Jan. 2014.
- [13] ISO/IEC IS 23009-1, "Information Technology – Dynamic adaptive streaming over HTTP (DASH)," Apr. 2012.
- [14] ITU-R Recommendation BT.500, "Methodology for the Subjective Assessment of the Quality of Television Pictures," *ITU-T*, Geneva, Switzerland, Jan. 2012.
- [15] ITU-T Recommendation P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," *ITU-T*, Geneva, Switzerland, Apr. 2008.
- [16] Q. Xu, Q. Huang, and Y. Yao, "Online Crowdsourcing Subjective Image Quality Assessment," in *Proc. of the 20th ACM international conference on Multimedia*, pp. 359-368, Nara, Japan, Oct. 2012.
- [17] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing Subjective Image Quality Evaluation," in *Proc. of 18th IEEE International Conference on Image Processing*, pp. 3097-3100, Brussels, Belgium, Sep. 2011.
- [18] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "QualityCrowd—A framework for crowd-based quality evaluation," in *Proc. Picture Coding Symposium*, pp. 245-248, Krakow, Poland, May. 2012.
- [19] T. Hossfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best Practices for QoE Crowdttesting: QoE Assessment with Crowdsourcing," *IEEE Trans. on Multimedia*, vol. 16, no. 2, pp. 541-558, Feb. 2014.
- [20] B. Gardlo, M. Ries, T. Hossfeld, and R. Schatz, "The impact of crowdsourcing platform choice on experimental results," in *Proc. International Workshop on Quality of Multimedia Experience*, pp. 35-36, Yarra Valley, Australia, Jul. 2012.
- [21] A. Takahashi, D. Hands, and V. Barriac, "Standardization activities in the ITU for a QoE assessment of IPTV," *IEEE Communications Magazine*, vol. 46, no. 2, pp. 78-84, Feb. 2008.



- [22] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," in *Proc. Int. Conference of Communications*, pp. 110-114, Beijing, China, May 2008.
- [23] M. Garcia and A. Raake, "Impairment-factor-based audio-visual quality model for IPTV," in *Proc. International Workshop on Quality Multimedia Experience*, pp. 1-6, California, US., Jul. 2009.
- [24] M. Martinez, M. Lopez, P. Pinol, M. Malumbres, and J. Oliver, "Study of Objective Quality Assessment Metrics for Video Codec Design and Evaluation," in *Proc. IEEE International Symposium on Multimedia*, pp. 517-524, California, US., Dec. 2006.
- [25] B. Ciubotaru, G.-M. Muntean, and G. Ghinea, "Objective assessment of region of interest-aware adaptive multimedia streaming quality," *IEEE Trans. on Broadcasting*, vol. 55, no. 2, pp. 202-212, Jun. 2009.
- [26] 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, pp. 76-78, Jan. 2011.
- [27] R. Mok, E. Chan, and R. Chang, "Measuring the Quality of Experience of HTTP Video Streaming," in *Proc. IEEE International Symposium on Integrated Network Management*, pp. 485-492, Dublin, Ireland, May 2011.
- [28] D. Rodriguez, J. Abrahão, D. Begazo, R. Lopes, and G. Bressan, "Quality metric to assess video streaming service over TCP considering temporal location of pauses," *IEEE Trans. on Consumer Electron.*, vol. 58, no. 3, pp. 985-992, Aug. 2012.
- [29] Q. Dai and R. Lehnert, "Impact of Packet Loss on the Perceived Video Quality," in *Proc. International Conference. on Evolving Internet*, pp. 206-209, Valencia, Spain, Sept. 2010.
- [30] M. Pinson and S. Wolf, "Application of the NTIA General Video Quality Metric VQM to HDTV quality monitoring," in *Proc. International Workshop Video Processing Quality Metrics for Consumer Electronics*, Chandler, US., Jan. 2007.
- [31] O. Sugimoto, S. Naito, S. Sakazawa, and A. Koike, "Objective perceptual picture quality measurement method for high-definition video based on full reference framework," in *Proc. International Society Optical Engineering*, vol. 7242, pp. 1-9, California, US, Jan. 2009.
- [32] J. Okamoto, K. Watanabe, A. Hondai, M. Uchida, and S. Hangai, "HDTV objective video quality assessment method applying fuzzy measure," in *Proc. International Workshop on Quality Multimedia Experience*, pp. 168-173, California, US, Jul. 2009.
- [33] VQEG, "Report on the validation of video quality models for high definition video content," *Tech. Report of Video Quality Experts Group*, Jun. 2010.
- [34] Z. Wang, A. Bovik, and H. Sheikh, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, No. 4, Apr 2004.
- [35] 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.
- [36] 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.
- [37] J. You, A. Perkins, M. Gabbouj, and M. M. Hannuksela, "Perceptual quality assessment based on visual attention analysis," in *Proc. International Conference on Multimedia*, pp. 561-564, Beijing, China, May 2009.
- [38] A. K. Moorthy 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.
- [39] ITU-T Recommendation J.1443, "User Requirements for Objective Perceptual Video Quality Measurements in Digital Cable Television," *ITU-T*, Geneva, May 2000.
- [40] ITU-T Recommendation P.10/G.100, "Appendix I - Definition of Quality of Experience (QoE)," *ITU-T*, Jan. 2007.
- [41] A. Raake and S. Egger, "Quality and Quality of Experience," in *Quality of Experience*, S. Moeller and A. Raake, Eds. London: Springer, 2014, pp. 11-34.
- [42] K. Wang, M. Barkowsky, K. Brunnstrom, M. Cousseau, and P. Le-Callet, "Perceived 3D TV transmission Quality Assessment: multi-laboratory results using absolute category rating on quality of experience scale," *IEEE Trans. on Broadcasting*, vol. 58, no. 4, pp. 544-557, Jun. 2012.
- [43] T. Zinner, O. Hohlfeld, O. Abboud, and T. Hohlfeld, "Impact of frame rate and resolution on objective QoE metrics," in *Proc. International Workshop on Quality of Multimedia Experience*, pp. 29-34, Trondheim, Norway, Jun. 2010.
- [44] B. Rainer, M. Waltl, E. Cheng, I. Burnett, and H. Hellwagner, "Investigating the impact of sensory effects on the QoE and emotional response in web videos," in *Proc. International Workshop on Quality of Multimedia Experience*, pp. 284-289, Yarra Valley, Australia Jul. 2012.
- [45] A. Moorthy, L. Choi, A. Bovik, and G. Veciana, "Video quality assessment on mobile devices: subjective, behavioral and objective studies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652-671, Oct. 2012.
- [46] D. Rodriguez and M. Arjona, "VoIP Quality Improvement with a Rate-determination Algorithm," in *Proc. 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, the M.S. degree (2009) and PhD degree (2013) from the University of São Paulo - USP. He is currently an Adjunct Professor at Department of Computer Science, Federal University of Lavras, Brazil. He has a solid knowledge in Telecommunication Systems and Computer Science based on 13 years of professional experience in important companies. His research interest includes QoS and QoE in multimedia services, Digital TV, IPTV and telecommunication networks.



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



**Eduardo Costa Alfaia** received his B.S. degree in Systems Analysis from FUCAPI, Brazil, and his M.S. degree from the Federal University of Amazonas (2011). He is a Ph.D. student at University of Brescia, Italy; and a fellow from the CAPES Foundation, Ministry of Education of Brazil. His current research interests include computer networks, multimedia services, recommendation systems, stream processing and big data.



**Julia Issy Abrahão** got her PhD degree in Ergonomics from the Conservatoire National des Arts etMetiers (1986) in Paris. Her Post-doc was obtained from ÉcolePratique des Hautes Études and Université Paris V. Nowadays she is an associate researcher at the 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.



**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 multimedia services, 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.

