Towards a General Parametric Model for Perceptual Video Quality Estimation

José Joskowicz, *IEEE Senior Member,* Rafael Sotelo, *IEEE Senior Member,* and J. Carlos Lopez Ardao, *IEEE Member*

***Abstract*—During the last few years, different parametric models were proposed for video quality estimation. Each model uses different parameters as inputs, such as bit rate, frame rate and percentage of packet loss, and each model was designed and tested by their authors for a particular codec, display resolution and/or application. This paper presents a review of the parametric models published by ten different groups of authors. Each model is briefly described, and the relevant parametric formulas are presented. The performance of each model is evaluated and contrasted to the other models, using a common video clips set, in different coding and transmission scenarios. Based on the results, a new and more general parametric model is presented, which takes into account bit rate, frame rate, display resolution, video content and the percentage of packet loss.**

***Index Terms*—Video perceptual quality, video quality parametric models, video signal processing**

I. Introduction

In RECENT years different evaluations and standardized efforts have been made, and are currently ongoing, in order to obtain objective models and algorithms to predict the perceived video quality in different scenarios.

The video quality models can be classified into FR (Full Reference), RR (Reduced Reference) and NR (No Reference) models [1]. In the first class, FR models, the original and the degraded video sequences are directly compared. In the RR models, some reduced information about the original video is needed, and is used along with the degraded video in order to estimate the perceived video quality. NR models are based only on the degraded video in order to make an estimation of the perceived video quality.

Based on the work of VQEG (Video Quality Experts Group) and other contributions, ITU-T has standardized some FR and RR models. Among them the Recommendation ITU-T J.144 [2] in 2004, the Recommendations ITU-T J.247 [3] and ITU-T J.246 [4] in 2008 and the Recommendation ITU-T J.249 in 2010 [5].

Manuscript received June 13, 2012; revised July 25, 2013; accepted July 26, 2013.

J. Joskowicz is with the Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, (e-mail: josej@fing.edu.uy).

R. Sotelo is with the Facultad de Ingeniería, Universidad de Montevideo, Montevideo, Uruguay, (e-mail: rsotelo@um.edu.uy).

J. C. L. Ardao is with ETSE Telecomunicación, Campus Universitario, Vigo 36310, Spain (e-mail: jardao@det.uvigo.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2013.2277951

Parametric models predict the perceived video quality based on a reduced set of parameters that are related to the encoding process, video content and/or transmission process (i.e. network information). These models typically present a mathematical formula, representing the estimation of the perceived video quality as a function of different parameters. Parametric models are easy to implement since there is no need to full access to the original video source. They may be applied to network design, network assessments and/or to real time monitoring. The quality estimation is easily computed as the result of a direct mathematical formula.

Many different parametric models have been proposed, with different scopes and applicable to different scenarios, and a parametric model has been standardized in the Recommendation ITU-T G.1070 in 2007 [6]. Each of the proposed parametric models has been evaluated by their authors. However, they usually study them in a particular use case. Nevertheless, a *general* parametric model that would apply for a wide range of applications, encoding parameters and transmission scenarios has not been developed yet.

In this paper we present a review of parametric models published in the last years by ten different groups of authors. The model's parameters and performance are evaluated and compared. The strengths and weaknesses of each model are remarked and are employed towards the development of a *general* parametric model for video quality estimation.

The paper is organized as follows: Section II describes main published parametric models. In Section III the performance of each model is presented. In Section IV the results of the previous section are analyzed, and based on them, a more general model is proposed. Section V summarizes the results and main contributions.

II. Parametric Models

Since this work aims to contribute towards a general parametric model, this section presents relevant parametric models proposed in last years. Each one is just briefly described, while its parametric formula is thoroughly detailed.

The presentation order is arbitrary and does not respond to any order of relevance.

A. Kazuhisa Yamagishi et al.: ITU-T G.1070 Model

ITU-T has published a model for predicting the video quality in video telephony applications, based on measurable parameters of the encoding process and the IP transport network. The Recommendation ITU-T G.1070 [6] describes a computational model for point-to-point interactive videophone applications over IP networks. The model is similar in form to the E-Model (Recommendation ITU-T G.107 [7]) and is based on the work performed by K. Yamagishi *et al.* [8], [9]. The model consists of three functions, one for video quality estimation ($\boldsymbol{V_q}$), another for audio quality estimation ($\boldsymbol{S_q}$) and the last for the overall multimedia quality estimation ($\boldsymbol{MM_q}$). Audio quality estimation is based on a simplification from ITU-T G.107 model. The video quality estimation function was patented in [10], and is performed according to Equation (1).

$$V_q = 1 + I_c I_t \tag{1}$$

where $\boldsymbol{V_q}$ is the estimation for MOS (Mean Opinion Score) or MOS_p (Predicted MOS), $\boldsymbol{I_c}$ is the video quality estimation determined by the encoding process and $\boldsymbol{I_t}$ is the video quality estimation determined by the transmission process. $\boldsymbol{I_c}$ depends on bit rate $\boldsymbol{b}$ and frame rate $\boldsymbol{f}$, according to equations (2) to (5).

$$I_c = I_o e^{-\frac{(\ln(f)-\ln(f_O))^2}{2D_{Fr}^2}} \tag{2}$$

$$f_o = v_1 + v_2 b \tag{3}$$

$$D_{Fr} = v_6 + v_7 b \tag{4}$$

$$I_o = v_3 \left(1 - \frac{1}{1 + \left(\frac{b}{v_4}\right)^{v_5}}\right) \tag{5}$$

$\boldsymbol{I_t}$ depends on bit rate and frame rate and the percentage of packet loss, according to equations (6) and (7).

$$I_t = e^{-\frac{p}{D_{Pplv}}} \tag{6}$$

$$D_{Pplv} = v_{10} + v_{11} e^{-\frac{f}{v_8}} + v_{12} e^{-\frac{b}{v_9}} \tag{7}$$

In these equations, $\boldsymbol{b}$ is the bit rate, $\boldsymbol{f}$ is the frame rate, $\boldsymbol{p}$ is the percentage of packet loss, and $\boldsymbol{v_1}$ to $\boldsymbol{v_{12}}$ are coefficients that must be calculated for each codec and display size or resolution. In this model, video content is not taken into account. The Recommendation states that the model handles videos whose size are between VGA (Video Graphics Array, 640×480 pixels) and QQVGA (Quarter Quarter VGA, 160×120 pixels), but provisional values for the coefficients are provided only for MPEG-4 in QVGA (Quarter VGA, 320×240 pixels) and QQVGA video formats. In [11] a new set of values for the coefficients are proposed for the MPEG-2 codec.

A similar model, for HDTV (High Definition TV, 1920×1080 pixels), was proposed by K. Yamagishi and T Hayashi in [12].

B. Fenghua You et al: T-Model

The ITU-T G.1070 model takes into account the packet loss, assuming a random loss distribution, but does not take into account a possible burst model for the packet loss pattern. Fenghua You *et al.* [13] have proposed an extension to the ITU-T G1070 model (the "T-Model"), according to equations (8) and (9).

$$I_t = e^{-\frac{p}{B_{Pplv} D_{Pplv}}} \tag{8}$$

$$B_{Pplv} = 1 + \alpha \frac{Den_{burst}}{N_{BP}} + \beta \frac{Den_{burst} D_{burst}}{Loss} \tag{9}$$

where $\boldsymbol{Den_{burst}}$ is the density of burst, $\boldsymbol{D_{burst}}$ is burst duration, and $\boldsymbol{Loss}$ is the total loss including both burst and gap loss. $\boldsymbol{N_{BP}}$ is the number of burst periods. Coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are dependent on codec, distortion concealment, and other factors related to media content.

The authors made subjective tests using MPEG-2 in HD with three video clips and they conclude that the proposed T-model achieves better accuracy than ITU-T G.1070 video model under burst loss conditions.

C. A. Raake et al.: T-V Model

In [14] A. Raake *et al*. have presented the "T-V Model", a parametric model for video quality estimation for SDTV (Standard Definition TV, 720×576 pixels) and HDTV. The model has a similar structure than the ITU-T G.107 E-Model, and was patented in [15]. Video quality estimation is performed according to equations (10) to (12).

$$V_q = Q_o - I_c - I_t \tag{10}$$

$$I_c = a_3 - a_1 e^{-a_2 b} \tag{11}$$

$$I_t = (b_o - I_c) \frac{p}{b_1 + p} \tag{12}$$

where $\boldsymbol{Q_o}$ is the maximum achievable quality (5 in the typical MOS scale), $\boldsymbol{b}$ is the bit rate, $\boldsymbol{p}$ is the percentage of packet loss and $\boldsymbol{a_1}$-$\boldsymbol{a_3}$, $\boldsymbol{b_o}$-$\boldsymbol{b_1}$ are coefficients that must be calculated for each codec and display size. Video content is not taken into account in this model.

The same authors, in [16], have made an analysis of the influence of video content. A preliminary analysis of the spatial and temporal complexity is presented and how these parameters influence the video quality, but only qualitative results were presented. In [17] an extension to the model is done, for IPTV in HD resolution, in order to take into account video content. The new model applies only to the degradation introduced in the encoding process, according to equation (13).

$$I_c = a_3 + a_6 MV_1 - a_1 e^{-a_2 b + a_4 QP_1} \tag{13}$$

where the new parameters are $\boldsymbol{MV_1}$ (the average of the standard deviation of the horizontal components of the Motion Vectors) and $\boldsymbol{QP_1}$ (Quantization Parameter per macro-block averaged over each I-frame), and $\boldsymbol{a_4}$-$\boldsymbol{a_6}$ are new coefficients.

In [18], Raake *et al.* presented a modification to the transmission impairment $\boldsymbol{I_t}$, based on the evaluation of the visibility of each lost packet. A "visibility classifier" module is described, and two parameters are extracted: the "estimated error" ($\boldsymbol{d_{mb}}$ the induced distortion, in terms of MSE, of the corrupted macroblocks which were noticeable in the frame where the loss occurred) and the "error propagation" ($\boldsymbol{d_{prop}}$ total number of impaired pixels due to error propagation), corresponding only to the packet classified as "visible". The resulting $\boldsymbol{I_t}$ formula is described in equation (14).

$$I_t = a_4 d_{mb} + d_{prop}^{a_5} + a_6 \tag{14}$$

D. H. Koumaras et al.: MPQoS Model

In [19] H. Koumaras *et al*. have presented the MPQoS (Mean Perceived Quality of Service) model. This model was designed for MPEG-4, in CIF (Common Intermediate Format, 352×288 pixels) and QCIF (Quarter CIF, 176×144 pixels) display sizes for multimedia applications. According to this model, the video quality can be estimated as

$$V_q = I_c = [PQ_H - PQ_L](1 - e^{-\alpha(b - BR_L)}) + PQ_L - 1 \tag{15}$$

where $\boldsymbol{b}$ is the bit rate and $\boldsymbol{PQ_H}$, $\boldsymbol{PQ_L}$, $\boldsymbol{BR_L}$ and $\boldsymbol{\alpha}$ are the four model coefficients. In this work, video quality was evaluated using the MPQoS metric, based on the PQM Picture Quality Metric proposed in [20]. According to the authors, the model coefficient can be derived using only one parameter $\boldsymbol{x}$ that depends on video content, according to equations (16). The impairments due to transmission factors ($\boldsymbol{I_t}$) were not modeled.

$$\alpha = f_1(x),\ BR_L = f_2(x),\ PQ_H = f_3(x) \tag{16}$$

The authors did not describe quantitatively how can be $\boldsymbol{x}$ derived from video content. A method was proposed to obtain the value of $\boldsymbol{x}$ based on having the video clip under evaluation coded at a very high bit rate.

In [21] we have shown that the "T-V Model" represented in Equation (11) and the MPQoS model represented in Equation (15) are equivalent, using the coefficients relations detailed in Equations (17).

$$a_1 = (PQ_H - PQ_L)e^{\alpha BR_L},\ a_2 = \alpha,\ a_3 = PQ_H - 1 \tag{17}$$

E. M. Ries et al. Model

In [22] M. Ries *et al*. have proposed a model for video quality estimation according to Equation (18).

$$V_q = I_c = A + Bb + \frac{C}{b} + Df + \frac{E}{f} \tag{18}$$

where $\boldsymbol{b}$ is the bit rate, $\boldsymbol{f}$ is the frame rate, and $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$ are the model coefficients. The authors have proposed to classify the video clips according to the video content, and for each class, a different set of coefficients are used. The authors show an algorithm to determine the content type and to make a classification into five classes. The degradation introduced in the transmission process is not evaluated. The $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$ model coefficients are calculated for H.264 with frame rates between 5 fps and 15 fps and bit rates between 24 kb/s and 105 kb/s.

F. J. Gustafsson et al. Model

J. Gustafsson *et al*. have proposed in [23] a model that takes into account the combined effects of packet loss and buffering. During the buffering time the image freezes, producing an annoying effect that affect the perceived quality. The model computes the video quality estimation based on the MOS for the original video clip, the buffer size in the receiver, the re-buffering time during reproduction and the packet loss in the network, and was evaluated for MPEG-4 in QCIF display size with bit rates up to 256 kb/s. The model details were patented in [24], and are presented in Equations (19) to (22)

$$V_q = 1 + (I_c - 1)I_t - I_b \tag{19}$$

$$I_c = c_0 - c_1 e^{-\lambda b} \tag{20}$$

$$I_t = k\frac{p_u - p_m}{p_u - p_l} \tag{21}$$

$$I_b = C_0 + C_1 InitP + C_2 BufP + C_3 BufF \tag{22}$$

where $\boldsymbol{b}$ is the bit rate, $\boldsymbol{p_u}$ and $\boldsymbol{p_l}$ are the upper and lower packet loss rate limit respectively, $\boldsymbol{p_m}$ is the average packet loss rate of the current logging window, $\boldsymbol{InitP}$ is the initial buffer time, $\boldsymbol{BufP}$ is the re-buffering time, $\boldsymbol{BufF}$ is the re-buffering events per minute and $\boldsymbol{k, c_o, c_1}$, λ, $\boldsymbol{C_0}$, $\boldsymbol{C_1}$, $\boldsymbol{C_2}$, $\boldsymbol{C_3}$ are the model coefficients. Video content was not taken into account in this model.

G. A. Khan et al. Model

In [25] A. Khan *et al*. have proposed a model for video quality estimation according to Equations (23) to (25)

$$V_q = I_c I_t \tag{23}$$

$$I_c = a_1 + a_2 f + a_3 \ln(b) \tag{24}$$

$$I_t = \frac{1}{1 + a_4 p + a_5 p^2} \tag{25}$$

where $\boldsymbol{b}$ is the bit rate, $\boldsymbol{f}$ is the frame rate, $\boldsymbol{p}$ is the packet loss, and $\boldsymbol{a_1}$-$\boldsymbol{a_5}$ are the model coefficients. The authors have developed the model using five video contents coded in H.264 in QCIF display size. They proposed to classify the video clips in three categories: "Slight Movement", "Gentle Walking" and "Rapid Movement".

The model was tested by the authors in QCIF display size, with frame rates between 10 fps and 30 fps, bit rates between 18 kb/s and 512 kb/s and with packet loss between 1% and 20%. The results were compared using the PSNR metric. No subjective tests were performed.

The same authors in [26] and [27] have presented a similar model, according to equations (26) to (28). This model was developed using five video contents coded in H.264 in QCIF display resolution.

$$V_q = a_1 + I_c I_t \tag{26}$$

$$I_c = a_2 e^{f} + a_3 \ln(b) + CT\,(a_4 + a_5 \ln(b)) \tag{27}$$

$$I_t = \frac{1}{1 + \left(a_6 p + a_7 p^2\right) a_8 B} \tag{28}$$

where the new parameters are $\boldsymbol{CT}$ (related to the Content Type) and $\boldsymbol{B}$ (Burst Length), and $\boldsymbol{a_1}$-$\boldsymbol{a_8}$ are the model coefficients.

H. Quan Huynh-Thu et al. Model

Quan Huynh-Thu *et al.* [28] have proposed a model of the impact of frame rate decimation, according to equation (29)

$$V_q = a_1 + \frac{a_2 - a_1}{1 + e^{a_3 f + a_4}} \tag{29}$$

where $\boldsymbol{f}$ is the frame rate, and $\boldsymbol{a_1} - \boldsymbol{a_4}$ are the model coefficients. The model was designed using seven video clips in QCIF display size, with frame rates between 2.5 fps and 30 fps. The authors analyze the relation between the video quality and video content, using the motion vectors as the estimation for the video motion content. But the results were not modeled in a parametric formula.

I. Yen-Fu Ou et al. Model

In [29] Yen-Fu Ou *et al*. have presented a model of the effect of frame rate, according to equation (30)

$$V_q = V_{q_{\max}} \frac{1 - e^{-c\frac{f}{f_{\max}}}}{1 - e^{-c}} \tag{30}$$

where $\boldsymbol{V_{qmax}}$ is the video quality obtained at the maximum frame rate $\boldsymbol{f_{max}}$ (30 fps in this case), $\boldsymbol{f}$ is the frame rate and $\boldsymbol{c}$ is the model coefficient. The model was derived using six video clips in CIF and QCIF display sizes with no noticeable degradations due to the encoding process (i.e. high bit rates). The frame rates used were between 6 fps and 30 fps. The authors state that the $\boldsymbol{c}$ coefficient depends on the video content, however an explicit formula for deriving $\boldsymbol{c}$ from video content was not presented.

$\boldsymbol{V_{qmax}}$ is modeled, in the referred paper, as

$$V_{q\max} = 1.04\left(1 - \frac{1}{1 + e^{0.34(PSNR - s)}}\right) \tag{31}$$

where PSNR is the non-perceptual Peak Signal to Noise Ratio, and $\boldsymbol{s}$ is a coefficient that depends on video content. The authors did not explain how to derive $\boldsymbol{s}$ from the video content.

J. Jose Joskowicz et al. Model

In [30] Jose Joskowicz *et al*. have proposed a model that combines the effects of frame rate, bit rate, display size and video content, as expressed in equations (32) to (34).

$$V_q = 1 + I_c \tag{32}$$

$$I_c = v_3 \left(1 - \frac{1}{1 + \left(\frac{ab}{v_4}\right)^{v_5}}\right) \tag{33}$$

$$\begin{aligned} v_3 &= 4 + 4(f_{\max} - f)(k_1 s + k_2 e^{-k_3(f_{\max} - f)ab}) \\ v_4 &= c_1 s^{c_2} + c_3 \\ v_5 &= c_4 s^{c_5} + c_6 \end{aligned} \tag{34}$$

where $\boldsymbol{b}$ is the bit rate, $\boldsymbol{f}$ is the frame rate, $\boldsymbol{f_{max}}$ is 25 fps, $\boldsymbol{a}$ is a constant that depends on display size, $\boldsymbol{s}$ is the average SAD (Sum of Average Differences) per pixel and $\boldsymbol{c_1}$-$\boldsymbol{c_6}$ and $\boldsymbol{k_1}$-$\boldsymbol{k_3}$ are the model coefficients. The model was derived using ten video clips, coded in H.264/AVC in VGA, CIF and QCIF display sizes at bit rates from 25 kb/s to 6 Mb/s and with frame rates from 5 fps to 25 fps.

The model takes into account the video content (using SAD as a characterization of the video content), but does not take into account the degradation introduced in the transmission process (i.e. packet loss).

III. Models Comparison

As can be seen from the previous section, many different parametric models have been proposed in last years. Each of the models were designed and/or tested at different conditions, taking into account specific parameters (i.e. bit rate, frame rate, video content, packet loss and so). A summary of the models is shown at Table I. In this section we will show and compare the results of the performance of the different models.

First, the performance of the models with respect to the $\boldsymbol{I_c}$ factor is presented. In this comparison, only the degradation introduced in the encoding process is evaluated (i.e., there are no packet loss or other degradations introduced in the transmission process). Then the $\boldsymbol{I_t}$ factor is evaluated for the models that includes the degradation of the transmission impairments, and another comparison is made.

The video clips detailed in Table II available in the VQEG web page [31] were used for the models comparison. Each clip was coded in H.264/AVC in bit rates from 100 kb/s to 6 Mb/s, in frame rates from 5 fps to 25 fps and in four different display sizes (SD, VGA, CIF, QCIF). Transmission impairment where performed with percentage of packet loss between 0% and 2%.

The best way to evaluate the performance of a model is contrasting the model results against the results obtained with subjective tests. Nevertheless, subjective tests are difficult to implement, and consume considerable time. In order to evaluate the performance of a model that takes into account different sets of parameters, including bit rate, frame rate, display size, video content and packet loss, a very large set of subjective test should be performed. For example, for the encoding process, using only 5 different video clips coded in 4 different bit rates, 4 different frame rates, and 4 different display sizes, will result in $5 \times 4 \times 4 \times 4 = 320$ video clips. In order to include the transmission process, different percentages of packet losses must be simulated for each of the former video clips. The impact of the same percentage of packet loss in a particular video clip will depend on the impacted frames (I, B or P). A packet loss in a I-frame will be propagated to other B- and P-frames, while a packet loss in a B-frame will not. In order to evaluate the *average* impact of a given percentage of packet loss in a specific video clip, many different loss patterns must be performed (with the same average percentage of packet loss). Considering 3 packet loss patterns for each percentage of packet loss, and 4 different percentages of packet loss, will lead to $320 \times 4 \times 3 = 3840$ video clips. Typically, no more than 30 video clips can be presented in the same subjective session with a duration of one hour, and only two people can make the subjective test at the same time (due to limitations in the angle and distance to the monitor). In order to obtain appropriate confidence intervals in the statistical results, more than 15 evaluators are needed for each video

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
MODELS COMPARISION

Ref	Author	Equations	Bit Rate	Frame Rate	Packet Loss	Packet Loss Burst	Packet Loss Visibility	Video Content – Metric	Disp Size	Re-Buff	# Coef	Tested Resolutions
A	K. Yamagishi ITU-T G.1070	2-7	Yes	Yes	Yes	No	No	No	No	No	12	VGA,QVGA; MPEG4
B	Fenghua You	8-9	Yes	Yes	Yes	Yes	No	No	No	No	2	HD, MPEG2
C1	A. Raake T-V Model	10-12	Yes	No	Yes	No	No	No	No	No	5	SD, HD; MPEG2, H.264
C2	A. Raake T-V Model	13	Yes	No	Yes	No	No	Yes - MV and QP	No	No	7	HD; H.264
C3	A. Raake	14	Yes	No	No	No	Yes	No	No	No	6	SD, HD; H.264
D	H. Koumaras: MPQoS Model	15	Yes	No	No	No	No	No	No	No	3	CIF, QCIF; MPEG4
E	M. Ries	18	Yes	Yes	No	No	No	Yes - Content Classes	No	No	5	CIF, QCIF, SIF; H.264
F	J. Gustafsson	19-22	?	No	Yes	No	No	No	No	Yes	?	QCIF; MPEG4
G1	A. Khan	23-25	Yes	Yes	Yes	No	No	Yes - Content Classes	No	No	5	QCIF; MPEG4
G2	A. Khan	26-28	Yes	Yes	Yes	Yes	No	Yes - ?	No	No	8	QCIF; H.264
H	Q. Huynh-Thu	29	No	Yes	No	No	No	No	No	No	4	QCIF
I	Yen-Fu Ou	30-31	No	Yes	No	No	No	No	No	No	1	CIF, QCIF
J	Jose Joskowicz	32-34	Yes	Yes	No	No	No	Yes – SAD	Yes	No	9	CIF, QCIF, VGA; H.264

clip [37]. So, for 3840 different degraded clips, more than 1000 subjective evaluation sessions should be performed.

In order to significantly reduce the number of subjective tests, we compared the models using two different sets of values.

The first one was obtained using a standard Reduced Reference model. In this case, many degraded video clips have been compared between a standard RR model and each of the parametric models evaluated. The second one was obtained with subjective tests, using a reduced set of video clips.

Full and Reduced Reference models have much more information than parametric models in order to predict the video quality (i.e. they have access to both, the original and the degraded video clip), and have been thoroughly evaluated by VQEG (Video Quality Experts Group) in different projects. We have used the LowBW (Low Bandwidth) Reduced Reference Model proposed by NTIA (National Telecommunications and Information Administration), standardized in 2010 in Recommendation ITU-T J.249 [5] and available in [32], as the VQM (Video Quality Metric) for the models performance comparison. This model is based on the “General Model” proposed by NTIA [33], standardized in Recommendation ITU-T J.144. The performance of the LowBW RR NTIA model for SD display size and 25/30 fps was well established in the VQEG RR/NR TV evaluations [34]. The result of these evaluations shows Pearson Correlation values from 0.82 to 0.88 between the subjective scores and the scores predicted from the NTIA model. The Pearson correlation metric evaluates the precision of the prediction. It varies from 0 to 1, where 1 indicates a direct relationship and 0 indicates no relationship at all. In this case, 0.82-0.88 indicates a high correlation between the values. Another evaluation of the NTIA model for SD display size was performed in [35], comparing the performance of different FR and RR models, and concluding that the NTIA model is one of the top performers.

TABLE II
VIDEO CLIPS

Src	Name	S-T Activity
2	Barcelona	High
3	Harp	Med
4	Moving graphic	Low
5	Canoa Valsesia	High
7	Fries	Med
9	Rugby	High
10	Mobile & Calendar	Med
13	Baloon-pops	High
14	New York 2	Low
16	Betes pas bêtes	Low
17	Le point	High
19	Football	High
21	Susie	Low
22	Tempete	Med

The NTIA model was originally designed and trained for small display sizes and low bitrates, and the overall performance of the model was presented by Stephen Wolf and Margaret H. Pinson in [36]. In the work presented in [30] we have made an independent evaluation of the NTIA model, for small display sizes and low frame rates, which shows that the Pearson Correlation value between the subjective scores and the LowBW RR NTIA Model is 0.91 and the RMSE (Root Mean Square Error) is 0.14. These results are better than the obtained in the VQEG model evaluation for SD display size and 25/30 fps. Using these results, the error of the NTIA model with respect to subjective scores can be estimated in +/- 15%.

For each video clip pair (original and degraded), the NTIA model provides a VQM, with values between 0 and 1 (0 when there are no perceived differences and 1 for maximum degradation), which corresponds to the DSCQS (Double Stimulus Continuous Quality Scale) [37] and can be directly associated with the DMOS (Difference Mean Opinion Scores). The DMOS values returned from the NTIA model can be related to the typical 5 points MOS using Equation (35).

$$MOS = 5 - 4DMOS \tag{35}$$

A. Comparison of Encoding Degradation Estimation

For the encoding degradation ($\boldsymbol{I_c}$), the performance of each model has been compared against the results of the NTIA model, using 10 different clips from Table II coded in CIF and VGA display sizes with bit rates from 100 kb/s to 6 Mb/s and frame rates from 5 to 25 fps. The clips used were sources 2, 3, 4, 5, 7, 9, 10, 14, 16, 17. These clips were selected in order to span over a wide range of contents. More than 470 encoded video clips were used for the comparison.

We have classified the clips in three different classes, according to the spatial and temporal (S-T) activity: "High", "Medium" and "Low" (see Table II). In the selected clips for this evaluation, four of them are categorized as having high S-T activity, three as medium S-T activity and three as low S-T activity. The best set of coefficients for each model was calculated. A different set of coefficients was calculated for each class of spatial and temporal activity (High, Medium, Low) for the models that include content classification.

The results are shown in Table III. The PC (Pearson Correlation), the RMSE (Root Mean Square Error), and the percentage of outlier points are presented (points outside the +/- 15% range). In this comparison, it is expected that a parametric model with "perfect" Pearson Correlation (i.e. PC = 1) with respect to the NTIA model will still have a +/-15% error with respect to subjective testes. On the other hand, all points that are in the +/- 15% region between the parametric model and the NTIA model have a precision equivalent to the standard RR model.

According to our results, the models that perform better for the encoded degradation ($\boldsymbol{I_c}$) take into account video content, bit rate and frame rate (models "J", "E" and "G1"). The best performance is obtained by model "J", with Pearson Correlation of 0.90 and only 8% of outlier points. This model uses the average SAD per pixel as an estimation of the coding complexity. The other two best performing models ("E" and "G1") make a video classification and use a different set of coefficients for each class.

In order to test the statistical significance of the performance improvement of model J compared to the other models we follow the same criteria adopted by VQEG in [34] for the PC. The sampling distribution of PC is not normally distributed, so "Fisher's z Transformation" is used which converts PC to the normally distributed variable z, according to (36).

TABLE III
$\boldsymbol{I_c}$ MODELS PERFORMANCE COMPARISON VS STANDARD RR MODEL

Ref.	Author	PC	RMSE	Outliers
A, B	K. Yamagishi	0.67	0.89	32%
C1, D	A. Raake / H. Koumaras	0.61	0.65	34%
E	M. Ries	0.79	0.45	16%
G1	A. Khan	0.76	0.54	25%
H	Quan Huynh-Thu	0.65	0.72	28%
I	Yen-Fu Ou	0.70	0.70	31%
J	J. Joskowicz	0.90	0.36	8%

$$z = 0.5 \cdot \ln\left(\frac{1 + PC}{1 - PC}\right) \tag{36}$$

The statistical significance test uses the H_0 hypothesis that assumes that there is no significant difference between correlation coefficients. The normally distributed $\boldsymbol{Z_{JK}}$ is determined for each comparison, according to (37).

$$Z_{JK} = \frac{z_j - z_k}{\sqrt{\frac{2}{N-3}}} \tag{37}$$

where z_j is the Fisher's z Transformation of the PC of model "J", z_k is the Fisher's z Transformation of the PC of each other model and $\boldsymbol{N}$ is the number of video clips used for the test. Since $\boldsymbol{N} > 30$, a normal distribution can be assumed, and thus $\boldsymbol{Z_{JK}} > 1.96$ implies that the model "J" is statically better than model $\boldsymbol{K}$ with a confidence of 95%. For every model represented in Table III, $\boldsymbol{Z_{JK}}$ is higher than 6, which imply a statistical confidence of almost 100% that Model "J" is better than the others.

A set of subjective tests were performed using 8 different video clips from Table II (sources 3, 9, 13, 14, 16, 19, 21 and 22), coded in 40 different combinations of display size (SD, VGA, CIF and QCIF), bit rates (500 kb/s to 3 Mb/s), frame rates (12.5 fps and 25 fps) and packet loss, resulting in 40 degraded video clips. The subjective tests were performed according to the Recommendation ITU-R BT.500-11 [37], using the DSCQS method. Each original – degraded video clip pair was evaluated by 19 subjects, between 23 and 64 years old, with an average of 28.7 years old. Twelve of them were males and seven were females. All of them are volunteers; some are professional workers and other are university students from different careers. From the 40 degraded video clips, 14 were produced without packet losses, i.e., with degradations produced only in the encoding process, and are the ones used for comparison in this section.

The comparison between the parametric models output and the subjective tests are presented in Table IV. Models "H" and "I" were not evaluated because these models require the video

TABLE IV
I_c MODELS PERFORMANCE COMPARISON VS SUBJECTIVE TESTS

Ref.	Author	PC	RMSE	Outliers
A, B	K. Yamagishi	0.62	0.44	14%
C1, D	A. Raake / H. Koumaras	0.57	0.48	14%
E	M. Ries	0.66	0.48	14%
G1	A. Khan	0.58	0.53	29%
J	J. Joskowicz	0.73	0.44	14%

quality of the decimated clip at the original frame rate, and the subjective tests were not designed to obtain this information.

Based on the subjective tests results, the model that performs better for the encoded degradation (I_c) is again model "J", with Pearson Correlation of 0.73 and 14% of outlier points.

In order to study the statistical significance of the results, the normal distribution is replaced by Student's t-distribution since the number of samples is only 14. In this case Z_{JK} should be higher than 2.15 to verify that the model "J" is statically better than other models with a confidence of 95%. This is not the case, and we cannot assure that model "J" is statistically better than the others using only the 14 clips evaluated in the subjective tests with degradations due to the encoding process.

B. Comparison of Transmission Degradation Estimation

For the transmission degradation (I_t), the performance of the models that include a "packet loss" parameter has been compared against the results of the NTIA model. The comparison was performed using six video clips from Table II (sources 3, 5, 9, 13, 14 and 22), encoded in VGA and CIF in more than 800 different configurations, with bit rates from 500 kb/s to 3 Mb/s, frame rates between 6.25 fps and 25 fps and percentage of packet loss from 0% to 2%, with random distribution. The clips were selected in order to have "low", "medium" and "high" spatial-temporal activity. The results are shown in Table V. The model that best performs for packet loss degradations is model "A". In this model, I_t depends not only on the percentage of packet loss, but also on the bit rate and frame rate. In the other two models, I_t only depends on the percentage of packet loss, but does not depend on bit rate or frame rate.

A set of subjective tests were performed with degradations in the transmission process (i.e., packet loss), as mentioned in the previous section. From the 40 degraded video clips used in the subjective tests, 26 were generated with packet loss, using 7 original video clips (sources 3, 9, 13, 14, 16, 21 and 22), coded in different configurations of display size (SD, VGA, CIF and QCIF), bit rate (500 kb/s to 3 Mb/s), frame rates (12.5 fps and 25 fps) and packet losses (from 0.2% to 2%).

The models that include a "packet loss" parameter where compared to the subjective scores and the results are presented in Table VI. In this comparison, again the model "A" has the best Pearson Correlation with the subjective tests, but model "C1" has the best RMSE and less number of outlier points.

TABLE V
I_t MODELS PERFORMANCE COMPARISON VS STANDARD RR MODEL

Ref.	Author	PC	RMSE	Outliers
A	K. Yamagishi	0.74	0.64	36%
C1	A. Raake	0.49	0.82	52%
G1	A. Khan	0.57	0.82	50%

TABLE VI
I_t MODELS PERFORMANCE COMPARISON VS SUBJECTIVE TESTS (PACKET LOSS FROM 0.2% TO 2%)

Ref.	Author	PC	RMSE	Outliers
A	K. Yamagishi	0.64	0.62	35%
C1	A. Raake	0.59	0.57	19%
G1	A. Khan	0.59	0.62	38%

IV. TOWARDS A GENERAL PARAMETRIC MODEL

Bit rate, frame rate, packet loss, display size, codec and video content are all relevant to make an estimation of the perceived quality for a given video clip. Each of the current proposed parametric models takes into account only a subset of these parameters. The models that perform better for the estimation of the encoding degradation (model "J") takes into account video content, bit rate and frame rate, and is the only model that explicitly includes the display size as a parameter. The model that performs better for the transmission degradation (model "A") takes into account the percentage of packet loss in combination with the bit rate and the frame rate.

None of the evaluated models explicitly takes into account video content in the transmission degradation. Nevertheless, the content affects the way that video quality is degraded by packet losses. As an example, Fig. 1 shows MOS values for clips src9 (Rugby) and src14 (New York) for different percentage of packet losses, calculated using the NTIA standard model. Both clips are coded in H.264/AVC, in VGA, at 3 Mb/s and 25 fps. As can be seen, the video quality of the clip src14 (New York) decays considerably faster with the percentage of packet losses than the clip src9 (Rugby).

In order to include the video content in the transmission factor (I_t), we have evaluated the video quality for six different video clips, coded in bit rates from 250 kb/s to 3 Mb/s, with frame rates from 6.25 fps to 25 fps, in two different display sizes (VGA and CIF) and with random packet losses between 0% and 2.5%. The selected video clips were src3 (Harp), src5 (Canoa), src9 (Rugby), src13 (Baloon-pops), src14 (New York) and src22 (Tempete). These sources span over a wide variety of video contents, with low, medium and high spatial-temporal activity. In total more than 800 different degraded video clips were generated. MOS values were

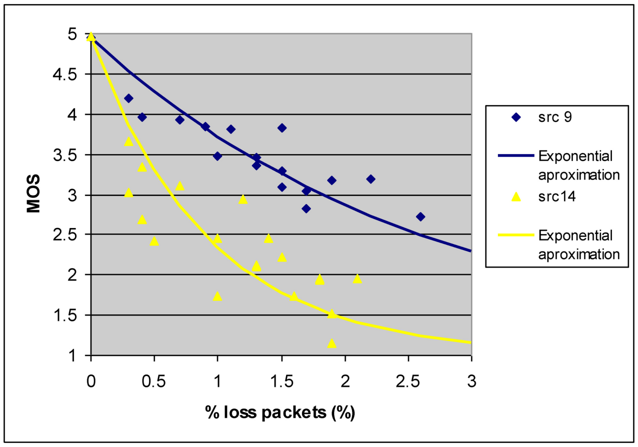

Fig. 1. MOS variation with respect to packet loss for two different video clips.

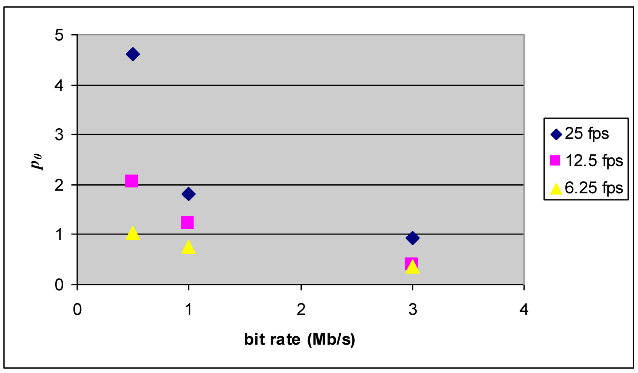

Fig. 2. p_0 with respect to bit rate, for different frame rates.

estimated with the NTIA standard model, and subjective tests were performed for a small subset of the video clips.

As a first step, we have studied how the video quality decreases for each clip with respect to the percentage of packet losses, leaving the bit rate and frame rate in fixed values. In average, for each clip, for a given bit rate and frame rate, the video quality can be modeled as expressed in equations (38) and (39)

$$V_q = 1 + I_{c_0} I_t \tag{38}$$

$$I_t = e^{-\frac{p}{p_0}} \tag{39}$$

were $\boldsymbol{I_{c0}}$ is the video quality for the clip obtained in the encoding process at the corresponding display size, bit rate and frame rate, $\boldsymbol{p}$ is the percentage of packet loss, and $\boldsymbol{p_0}$ is a constant, with different values for each clip, bit rate and frame rate. An example of this approximation is presented in the solid lines of Fig. 1. Equation (39) is the same as Equation (6) of model "A".

As a second step, the relation between $\boldsymbol{p_0}$ with respect to bit rate, frame rate and video content was evaluated. Fig. 2 shows how $\boldsymbol{p_0}$ varies with respect to the bit rate for different frame rates for the clip src22 ("Tempete") coded in H.264/AVC in VGA. For each frame rate, $\boldsymbol{p_0}$ decays with bit rate in a negative exponential form. Similarly, Fig. 3 shows how $\boldsymbol{p_0}$ varies with

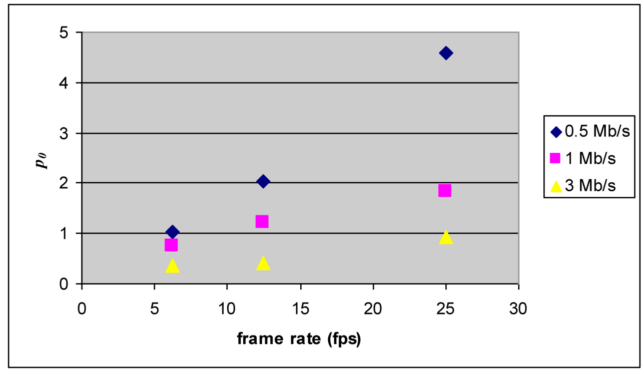

Fig. 3. p_0 with respect to frame rate, for different bit rates.

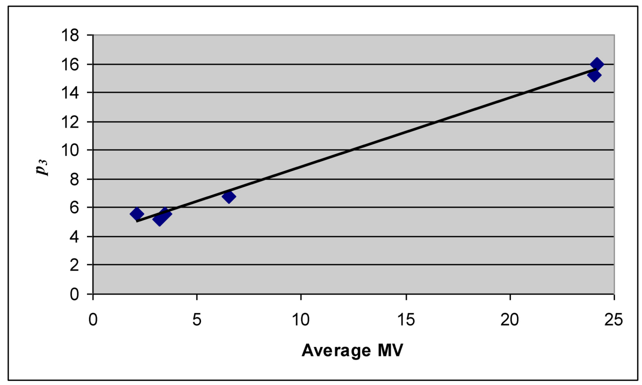

Fig. 4. p_3 with respect to the average amplitude of the Motion Vectors.

respect to the frame rate for different bit rates for the same clip. For each bit rate, $\boldsymbol{p_0}$ shows a linear relation with respect to the frame rate. Similar relations can be observed for different video clips and in different display sizes.

With these considerations, for each clip, $\boldsymbol{p_0}$ can be expressed as

$$p_0 = p_1 + p_2 f + p_3 e^{-\frac{b}{b_0}} \tag{40}$$

Where $\boldsymbol{b}$ is the bit rate, $\boldsymbol{f}$ is the frame rate, and $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, $\boldsymbol{p_3}$, $\boldsymbol{b_0}$ are constants for each clip, but may be different for different clips. Equation (40) is similar to the equation (7) proposed in model "A", but in equation (40) a linear relation with respect the frame rate is proposed instead of an exponential relation. This is based on the results obtained with more than 800 video clips used for this evaluation.

As stated in the earlier paragraphs, the coefficients of Equation (40) still depend on the video clip spatial and temporal content. We have found a strong relation between coefficients $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, $\boldsymbol{p_3}$ and the average value of the amplitude of Motion Vectors of the clip. Fig. 4 show the relation between $\boldsymbol{p_3}$ and the average value of the amplitude of Motion Vectors, for the six video clips used. Similar relations can be found for $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$.

According to these observations, the coefficients $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, $\boldsymbol{p_3}$ can be expressed as

$$\begin{aligned} p_1 &= c_7 m + c_8 \\ p_2 &= c_9 m + c_{10} \\ p_3 &= c_{11} m + c_{12} \end{aligned} \tag{41}$$

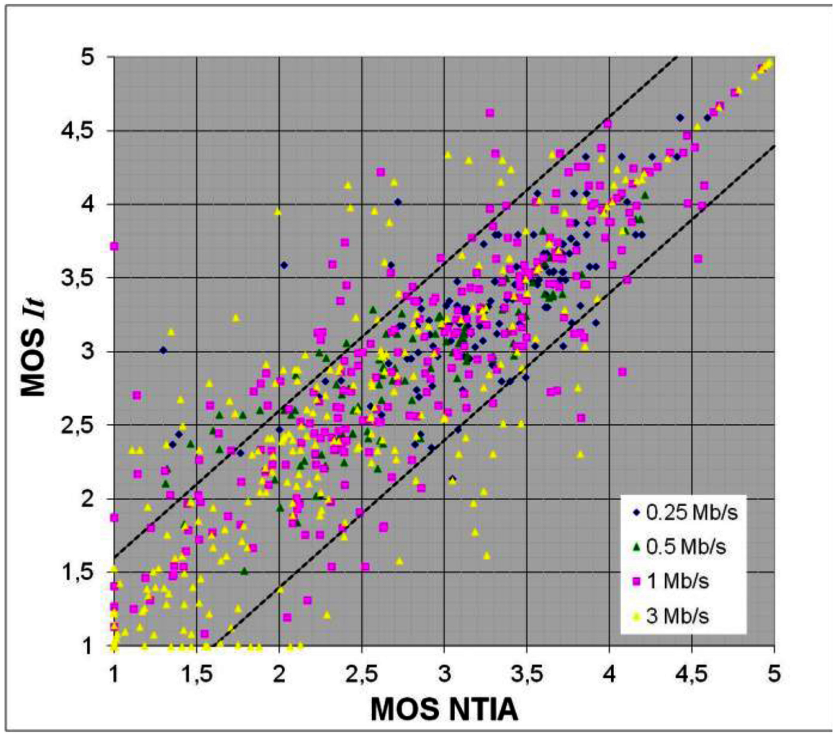

Fig. 5. I_t dispersion between J1 Model and NTIA model.

TABLE VII
J1 MODEL PERFORMANCE VS NTIA SCORES

Ref.	Author	PC	RMSE	Outliers
J1	J. Joskowicz	0.87	0.49	17%

TABLE VIII
J1 MODEL PERFORMANCE VS SUBJECTIVE SCORES

Ref.	Author	PC	RMSE	Outliers
J1	J. Joskowicz	0.89	0.47	20%

where $\boldsymbol{m}$ is the average amplitude of the Motion Vectors, calculated for each clip and c_7–c_{12} are constants. On the other hand, the coefficient b_0 does not show such a relation with respect to $\boldsymbol{m}$, or with respect to the Average SAD per pixel $\boldsymbol{s}$, and can be assumed as a constant for all the clips.

Using I_t as expressed in equations (39), (40) and (41) (let's call this Model "J1"), Table VII shows the obtained Pearson Correlation (PC), Root Mean Square Error (RMSE) and Outliers, comparing more than 800 video clips between the standard RR model and the parametric model "J1". The results are better than those obtained with model "A" (compare to Table V). Fig. 5 shows the dispersion between the NTIA standard model and the "J1" model for the $I_t$factor estimation. Each symbol represents video clips coded at different bitrates.

Again, we checked the statistical significance of the performance improvement of model "J1" compared to the other models depicted in Table V, using the same method. In this case, for every model represented in Table V, Z_{JK} is higher than 7, which imply a statistical confidence of almost 100% that model "J1" is better than the others.

According to these results, a more *general* parametric model for video quality estimation may be derived from a combination of Model "J" for the estimation of the encoding degradation and Model "J1" for the estimation of transmission degradation. This model takes into account video content, bit rate, frame rate, display size and percentage of packet loss as the relevant parameters for video quality estimation. Video content analysis in this model is based on the average SAD per pixel, as an estimation of the coding complexity (affecting the I_c factor) and the average of the amplitude of Motion Vectors, as this is relevant to the error propagation when packet loss occurs (affecting the I_t factor)

The overall performance of the Model "J-J1" was evaluated with subjective tests, using 40 video clips, coded in H.264/AVC, in different combinations of display sizes (SD, VGA, CIF and QCIF), bit rates from 500 kb/s to 3000 Mb/s, frame rates of 12.5 fps and 25 fps and percentage of packet loss from 0% to 2%. The results are presented in Table VIII, and the dispersion between the proposed model and the subjective scores are showed in Fig. 6.

In order to study the statistical significance of the results, we compared each result presented in Table VI against model "J-J1". We used the Student t-distribution with N = 26. The corresponding values of Z_{JK} for models "A", "C1" and "G1" are 1.95, 2.22 and 2.22 respectively. That implies that model "J-J1" has better performance than model "A" with a confidence of 90%, and with a confidence of 95% for the other two models.

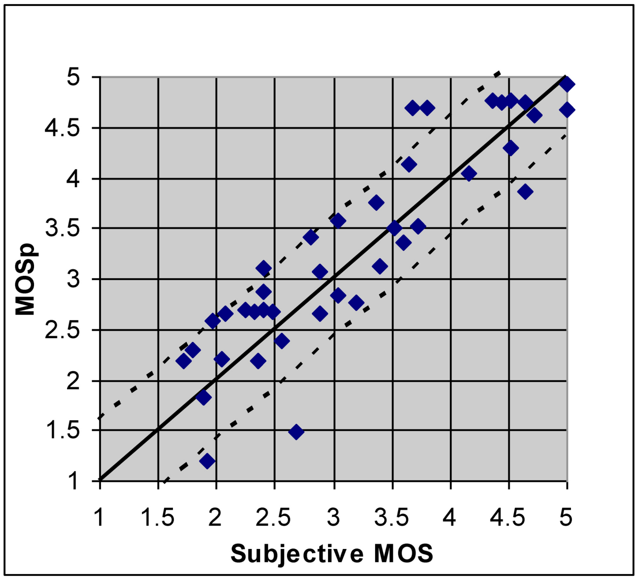

Fig. 6. Proposed model dispersion vs Subjective scores.

Other factors that were not evaluated may also affect the perceived quality. Parameters such as the available bandwidth (causing re-buffering), GOP size and structure, packet loss burstiness and concealment strategy, delay and jitter, video filters at receiver, codec specific configurations and display type, among others should be considered towards a more *general* model.

V. Conclusion

Parametric models for video quality estimation proposed by ten different groups of authors and organizations in the last years were presented and analyzed. A performance comparison was performed for the encoding and transmission impairments estimation of each model. A new method for the evaluation of the performance of each model has been proposed and used, consisting in the combination of the comparison of the parametric models with standard RR models using a large set of video clips and subjective tests using a reduced set of video clips.

According to the obtained results, it can be seen that the model that performs better for the encoding impairments estimation is the proposed by Jose Joskowicz *et al*. in [30], and the model that performs better for the transmission impairments estimation is the proposed in the Recommendation ITU-T G.1070 [6]. None of the evaluated models take into account the video content in the transmission impairments estimation. It has been shown that the video content affects the way that packet losses affects the perceived video quality, and a new model for the transmission impairments estimation has been proposed in this paper. Combining the model proposed in [30] for the encoding quality estimation with the new proposed model for the transmission quality estimation, a new overall model is presented, that takes into account video content, bit rate, frame rate, display size and percentage of packet loss as the relevant parameters for video quality estimation. In this model, video content analysis is based on the average SAD per pixel, as an estimation of the encoding complexity (affecting the I_c factor) and the average of the amplitude of Motion Vectors, as this is relevant to the error propagation when packet loss occurs (affecting the I_t factor). The new model performance has been compared to other proposed parametric models. It has shown a better performance for both, encoding and transmission degradations. Also, subjective tests show very good results, with a Pearson Correlation of 0.89.

Towards a more general model, other factors (such as GOP size and structure, packet loss concealment strategy, video filters at receiver, codec specific configurations and display type) should be explored, and may also be incorporated in a *general* parametric model for video quality estimation.

References

[1] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 660–668, Sep. 2008.

[2] *Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference*, Rec. ITU-T J.144, Feb. 2004.

[3] *Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference*, Rec. ITU-T J.247, Aug. 2008.

[4] *Perceptual Visual Quality Measurement Techniques for Multimedia Services Over Digital Cable Television Networks in ihe Presence of a Reduced Bandwidth Reference*, Rec. ITU-T J.246, Aug. 2008.

[5] *Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Reduced Reference*, Rec. ITU-T J.249, Jan. 2010.

[6] *Opinion Model for Video-Telephony Applications*, Rec. ITU-T G.1070, Apr. 2007.

[7] *The E-model, a computational model for use in transmission planning*, Rec. ITU-T G.107, Mar. 2005.

[8] K. Yamagishi and T. Hayashi, "Opinion model for estimating video quality of videophone services," in *Proc. IEEE Global Telecommun. Conf.*, 2006, pp. 1–5.

[9] T. Hayashi, K. Yamagishi, T. Tominaga, and A. Takahashi, "Multimedia quality integration function for videophone services," in *Proc. IEEE Global Telecommun. Conf.*, 2007, pp. 2735–2739.

[10] K. Yamagishi and T. Hayashi, "Video quality estimation apparatus, method, and program," U.S. Patent US2010/0284295 A1, Nov. 11, 2010.

[11] B. Belmudez and S. Moöller, "Extension of the G.1070 video quality function for the MPEG2 video codec," in *Proc. 2nd Int. Workshop QoMEX*, 2010.

[12] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," in *Proc. IEEE ICC*, 2008.

[13] F. You, W. Zhang, and J. Xiao, "Packet loss pattern and parametric video quality model for IPTV," in *Proc. IEEE/ACIS 8th Int. Conf. Comput. Inf. Sci.*, 2009.

[14] A. Raake, M. N. Garcia, S. Moller, J. Berger, F. Kling, P. List, and J. H. Johann, "T-V-model: Parameter-based prediction of IPTV quality," in *Proc. IEEE ICASSP*, 2008.

[15] A. Raake, M. N. Garcia, and S. Moller, "Audio-visual quality estimation," U.S. Patent US2009/0244289 A1, Oct. 1, 2009.

[16] M. N. Garcia, A. Raake, and P. List, "Towards content-related features for parametric video quality prediction of IPTV services," in *Proc. ICASSP*, 2008, pp. 757–760.

[17] M.-N. Garcia, R. Schleicher, and A. Raake, "Towards a content-based parametric video quality model for IPTV," in *Proc. 3rd Int. Workshop PQS*, 2010, pp. 20–25.

[18] S. Argyropoulos, A. Raake, M.-N Garcia, and P. List, "No-reference bit stream model for video quality assessment of H.264/AVC video based on packet loss visibility," in *Proc. IEEE ICASSP*, 2011, pp. 1169–1172.

[19] H. Koumaras, A. Kourtis, D. Martakos, and J. Lauterjung, "Quantified PQoS assessment based on fast estimation of the spatial and temporal activity level," *Multimedia Tools Appl.*, vol. 34, no. 3, pp. 355–374, Sep. 2007.

[20] J. Lauterjung, "Picture quality measurement," in *Proc. IBC*, 1998, pp. 413–417.

[21] J. Joskowicz and J. Carlos López Ardao, "A parametric model for perceptual video quality estimation," *Telecommun. Syst., Springer Spl. Issue Quality Exp. Issues Multimedia Provision*, vol. 49, no. 1, pp. 49–62, Jan. 2012.

[22] M. Ries, C. Crespi, O. Nemethova, and M. Rupp, "Content based video quality estimation for H.264/AVC video streaming," in *Proc. WCNC*, 2007, pp. 2668–2673.

[23] J. Gustafsson, G. Heikkilä, and M. Pettersson, "Measuring multimedia quality in mobile networks with an objective parametric model," in *Proc. IEEE ICIP*, 2008, pp. 405–408.

[24] J. Gustafsson, G. Heikkilä, M. Pettersson, and X. Tan, "Method of determining video quality," U.S. Patent 2010/0008241 A1, Jan. 14, 2010.

[25] A. Khan, L. Sun, and E. Ifeachor, "Content-based video quality prediction for MPEG4 video streaming over wireless networks," *J. Multimedia*, vol. 4, no. 4, pp. 228–239, Aug. 2009.

[26] H. Koumaras, F. Liberal, J. O. Fajardo, I. Taboada, J. Lauterjung, I. Gkika, J. Putty, M. Sidibe, I.-H. Mkwawa, A. Khan, M. Goudarzi, E. Jammeh, L. Sun, and E. Ifeachor, "ADAMANTIUM: Adaptative management of media distribution based on satisfaction oriented user modelling. D4.1–voice and video quality perceptual models," Information and Communication Technologies. Contract No. 214751.

[27] A. Khan, L. Sun, and E. Ifeachor, "QoE prediction model and its application in video quality adaptation over UMTS networks," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 431–442, Apr. 2012.

[28] Q. Huynh-Thu and M. Ghanbari, "Temporal aspect of perceived quality in mobile video broadcasting," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 641–651, Sep. 2008.

[29] Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 286–298, Mar. 2011

[30] J. Joskowicz and J. C. López Ardao, "Combining the effects of frame rate, bit rate, display size and video content in a parametric video quality model," in *Proc. LANC*, 2011, pp. 4–11.

[31] *VQEG phase I test sequences*. [Online]. Available: ftp://vqeg.its.bldrdoc.gov/SDTV/VQEG_PhaseI/TestSequences/Reference/

[32] *Video quality metric (VQM) software*. [Online]. Available: www.its.bldrdoc.gov/n3/video/vqmsoftware.htm

[33] M. H. 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.

[34] Video Quality Experts Group, Validation of Reduced-Reference and No-Reference Objective Models for Standard Definition Television, Phase 1, Final Rep., 2009.

[35] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[36] S. Wolf and M. H. Pinson, "Low bandwidth reduced reference video quality monitoring system," in *Proc. 1st Int. Workshop Video Process. Quality Metrics*, 2005.

[37] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Rec. ITU-R BT.500-11, Jun. 2002.

José Joskowicz (M'09–SM'10) is an Associate Professor at the Electric Engineering Institute at the Engineering School, University of the Republic of Uruguay, Uruguay. He is an Electronics Engineer specialized in Telecommunications from the University of the Republic of Uruguay, in 1995.

He received the Ph.D. degree in Telematics Engineering from the Unversity of Vigo, Vigo, Spain, in 2012.

He is currently a Consulting Engineer in Telecommunications areas.

His main research interests are in the area of quality of service and quality of experience in multimedia applications.

Rafael Sotelo (M'05–SM'13) is Vice Chair of Uruguay Section 2013–2015, and BTS member.

He is the Director of ICT Department at the Universidad de Montevideo, Uruguay, and an Associate Professor at the Universidad de la República, Uruguay. He became an Electrical Engineer in 1994 and an MBA in 1999. He received the Ph.D. degree in Telematics Engineering from the Universidad de Vigo, vigo, Spain, in 2010. He has been associated with the broadcasting and cable TV industry since 1991.

His main research interests are in the area of quality of experience in multimedia applications, transmission of DTV and recommender systems.

J. Carlos Lopez Ardao (M'13) is an Associate Professor in the Telematic Engineering Department at the University of Vigo, Vigo, Spain.

He received the Ph.D. degree in telecommunication engineering from the University of Vigo, in 1999.

His main research interests are in the area of performance evaluation, traffic engineering, quality of service, and multimedia applications in next-generation-networks.